MECHANISM AND ENERGETICS OF MEMBRANE TRANSPORTERS AND CHANNELS

BY

GIRAY ENKAVI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biophysics and Computational Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Emad Tajkhorshid, Chair
Professor Klaus Schulten
Professor Robert B. Gennis
Professor Satish K. Nair

Abstract

Living cells have evolved specialized transport proteins called membrane transporters and channels that catalyze exchange of materials across the cell membrane. Membrane transporters couple the active transport of their specific substrates against their electrochemical gradient. On the other hand, membrane channels facilitate passive diffusion of polar or charged molecules down their electrochemical gradient. We present here molecular dynamics (MD) investigation of a membrane transporter, glycerol-3 phosphate transporter (GlpT) and two membrane channels, urea transporter (UT) and aquaporin (Aqp1). Each simulations presented here provided a dynamical and atomistic picture of the protein of interest in a collaborative effort with an experimental lab.

Membrane transporters use various source of cellular energy, e.g., ATP binding and hydrolysis in primary active transporters, and pre-established electrochemical gradient of molecular species other than their substrate in the case of secondary active transporters. All membrane transporters use the widely-accepted "alternating-access mechanism", which ensures that the substrate is only accessible from one side of the membrane at a given time, and relies on complex protein conformational changes between outward-facing (OF) and inward-facing (IF) states, going through several intermediate states.

The first system that we investigated is the glycerol-3-phosphate transporter (GlpT), an antiporter member of the MFS. GlpT transports glycerol-3-phosphate (G3P) into the cell in exchange for inorganic phosphate (P_i). Major facilitator superfamily (MFS) is the largest superfamily of secondary active transporters and catalyze the transport of an enormous variety of small solute molecules across biological membranes. Individual MFS members, despite their architectural similarities, exhibit strict specificity toward

the substrates that they transport. The structural basis of this specificity, however, is poorly understood. Our collaborators, Da-Neng Wang Lab (New York University, NY) performed mutagenesis studies and transport assays, while we performed equilibrium simulations of wild-type GlpT and several of its mutant forms in membrane in the presence of all physiologically relevant substrates (P_i^-, P_i^{2-}, $G3P^-$, and $G3P^{2-}$) to characterize the determinants of substrate selectivity and conformational response of the protein to substrate binding. The positive electrostatic potential of the lumen of GlpT recruits substrate and drives binding. Only a few amino acid residues that line the transporter lumen act as specificity determinants. The phosphate moiety of P_i and G3P bind to a common binding site and residues involved solely in recognition of the glycerol moiety of G3P confers it a higher binding affinity. Furthermore, the simulations characterized the process and mechanism of substrate binding, and the protein's initial conformational response. All substrate-bound systems resulted in partial closing of the cytoplasmic half of GlpT. Extended simulations of substrate-bound systems also captured a water-conducting "channel-like" state. These states were also observed in several other transporters, suggesting that alternating-access mechanism tolerates transient states that are partially open to both sides of the membrane. We, later, obtained a model of the outward-facing (OF) state of GlpT using nonequilibrium molecular dynamics and calculated free energies to investigate the energetics associated with the transport cycle of GlpT.

The second system we report here is a membrane channel that facilitates passive diffusion of urea across the membrane, namely the urea transporter (UT). Urea is ubiquitously used as a nitrogen source by bacteria and a safe end product of protein catabolism. Due to its highly polar nature, urea relies on the UTs to permeate through the cell membrane. UTs are most frequently found in kidneys of mammals and allow rapid equilibration of urea between the urinary space and the hyperosmotic tissue fluid to prevent osmotic diuresis. Our collaborators, Ming Zhou lab (Columbia University, NY), crystallized structure of a mammalian UT (UT-B). UT-B is a homotrimer and each monomer contains a urea conduction pore with a narrow selectivity filter. We performed an extensive set of molecular dynamics simulations combined with free energy calculations to elucidate the

structural determinants of the selectivity in UT-B and the associated energetics. The orientation of the urea as its goes through the channel, as well as specific water-urea and protein-urea interactions determine the specificity of the channel. The free energy barrier at the selectivity filter appears to be approximately 5.0 kcal/mol. We, then, investigated the gas permeability of UT in collaboration with Walter F. Boron lab (Case Western University, Cleveland). Our free energy calculations along with the physiological experiments indicate that UTs can function as gas channels and identified the monomeric pores as the main conduction pathway for both water and NH_3. Our work characterized UTs as the third family of gas channels along with aquaporins (water channels), and Rh-associated glycoprotein (RhAG) (ammonia channels).

The other membrane channel system that we investigated is an aquaporin. Aquaporins are ubiquitous integral membrane channels that maintain water homeostasis of the cell by facilitating selective diffusion of water across the membrane while preventing proton diffusion. Two conserved regions located along the pore are responsible for the selectivity: the dual asparagine, proline, alanine (NPA) aquaporin signature motif, and the aromatic/arginine selectivity filter (SF). Recently, our collaborators, Richard Neutze lab (University of Gotenburg, Sweden), have crystallized a yeast aquaporin at 0.88 Å resolution, the highest resolution achieved to date for a membrane protein. The structure reveals a great deal of novel information on the structure of hydrogen-bonded network of water and protein side chains. To determine the dynamics and energetics of water diffusion along the channel, we performed molecular dynamics simulations of this impressively high quality crystal structure. The results show disruption of the water chain in both NPA and SF regions in this aquaporin, due to characteristic hydrogen-bonding patterns that dictate specific orientations to water molecules. The motion of water molecules is highly correlated on either side of the NPA region. The correlation, however, is lower at the NPA region, attesting yet another possible mechanism for this region to contribute to a barrier against proton transport. Besides, the NPA region appears as a barrier region with low occupancy for water, a feature not seen in other aquaporins. The correlated motion of adjacent water molecules along with their binary co-occupancies in the SF show that

water molecules move in pairs in this region. Specific hydrogen-bonding patterns in the SF region may also play a role in exclusion of hydronium (H_3O^+) and/or hydroxide ions (OH^-). These simulations have helped elucidate the dynamical basis of many intricate features revealed by this new structure.

Acknowledgments

I am profoundly grateful to my advisor, Emad Tajkhorshid for giving valuable guidance, support, and advice on all of the exciting projects in my research and helping me to develop this document. I would also like to thank all my collaborators at Da-Neng Wang Lab, Ming Zhou lab, Richard Neutze lab, and Walter F. Boron lab for providing insightful discussion about the projects and opportunity to publish together. I also appreciate the guidance and support of my committee members, Professor Klaus Schulten, Professor Robert B. Gennis, and Professor Satish K. Nair, and thank them for their time. I would finally like to thank all the current and past members of Emad Tajkhorshid lab, especially Zenmei Ohkubo and Mahmoud Moradi, for valuable support, discussions, and friendship.

Table of Contents

Chapter 1

Membrane Transport Proteins

Living cells rely on continuous exchange of diverse molecular species, e.g., nutrients, precursors, and reaction products, across the cellular membrane for their proper function [1]. Apart from small and/or highly hydrophobic molecules that can readily permeate lipid bilayers, for almost every chemical species, specialized integral membrane transport proteins have evolved that facilitate their traffic across the membrane in a highly selective manner. These transport proteins are called membrane transporters and channels depending on their mechanism of transport.

1.1 Membrane Transporters *

Membrane Transporters often facilitate the translocation of the substrates against their electrochemical gradients, a process which relies on delicate coupling mechanisms between substrate transport and various sources of cellular energy exploited by the transport mechanism. Membrane transporters couple various sources of cellular energy to vectorial translocation of their substrates, often against the chemical gradient. The required energy is provided by ATP, redox reactions, or photons in *primary* transporters, while secondary transporters couple substrate transport to co-transport (*symport* or *an-*

*This section is largely based on a review article in Journal of Computational and Theoretical Nanosciences [2]. Saher A. Shaikh and Po-Chao Wen and Giray Enkavi and Zhijian Huang and Emad Tajkhorshid "Capturing functional motions of membrane channels and transporters with molecular dynamics simulation". *Journal of Computational and Theoretical Nanosciences*, 7:2481–2500, **2010**.

tiport) of species with an already established gradient across the cell membrane. The energy-coupling mechanism relies on a complex set of molecular processes within the transporter protein engaging many of its structural elements, rendering the dynamics of membrane transporters highly relevant to their function. Relevant structural motions of membrane transporters span a wide spectrum of scales, ranging from localized side-chain conformational changes, to loop flipping motions, and up to extensive subdomain/domain structural transitions. Molecular Dynamics [3] simulation offers a method with sufficient temporal and spatial resolutions to characterize functionally relevant molecular events in proteins [4–12].

The transport cycle in membrane transporters is usually composed of an unknown number of steps, often resulting in very slow turnover of the process. These steps usually include protein conformational changes of different natures and magnitudes that are triggered by various molecular events, e.g., binding, unbinding, and stepwise translocation of the substrate and its cotransported ions, hydrolysis of ATP, and hydration and dehydration of various compartments within the transporter protein. Specific protein conformational changes coupled to these events are at the heart of the mechanisms employed for energy coupling and efficient transport by membrane transporters.

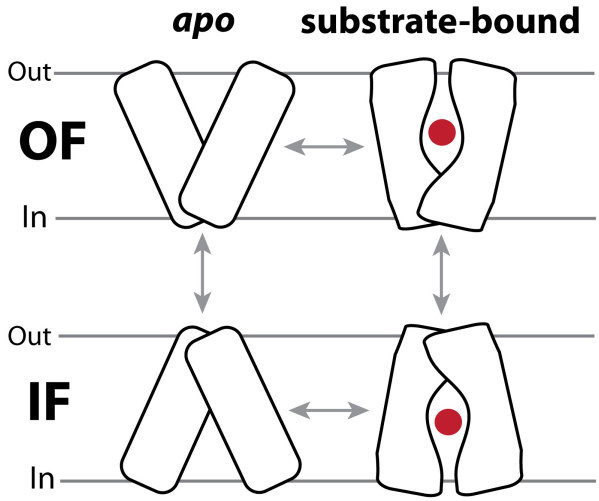

Figure 1.1: The alternating access model of membrane transporters, including two major states, outward facing (OF) and inward facing (IF) states, which can exist for both the *apo* and substrate-bound forms of the transporter. Substrate access from the two sides of the membrane is controlled by protein conformational changes.

The widely accepted general mechanistic model for transporters, termed the alternating-access model [13], proposes that the transporter protein switches substrate accessibility from one side of the membrane to the other, through undergoing structural transitions between outward-facing (OF) and inward-facing (IF) states, temporarily residing in several possible intermediate states (Fig. 1.1). Crystal structures for several of these proposed states [14–17] along with biochemical, kinetic and structural studies [18, 19] across various families of transporters have provided strong support for this mechanism. However, the availability of structures for multiple functional/conformational states of the same protein, or even within the same family is limited [20–31]. Even when multiple conformational states have been characterized for the same membrane transporter, how transitions between these states and how they are coupled to the energy-providing mechanism is largely unknown. Therefore, reconstructing the transport cycle and understanding the mechanism of energy coupling and substrate/ion transport rely on techniques that would yield a dynamical description of the process.

1.2 Membrane Channels

Membrane Channels, on the other hand, facilitates passive diffusion of ions (ion channels), water (aquaporins) or other solutes through the membrane down their electrochemical gradient. These polar/hydrophilic species either are membrane-impermeable or diffuse very slowly through the membrane. However, many physiologically processes require fast equilibration of these substances across the cell membrane, such as in the case of generation of action potentials in nerve cells, muscle contraction, etc. [32, 33]

Membrane channels are classified into four types based on their "gating" properties in response to stimuli: the voltage-gated channels which respond to changes in membrane potential; the ligand-gated channels, in which ligand binding triggers an allosteric conformational response to open/close the channel; the mechanically gated channels, which respond to mechanical stimuli such as membrane tension; and non-gated channels [32]. Other types of classifications of channels are generally based on the transported sub-

strate. In the case of ion channels, they are classified based on the type of the ion that they transport, and aquaporins are named after water.

Unlike membrane transporters that require large conformational changes to actively transport their substrate across the membrane, membrane channels, in general, do their job with minimal conformational change [32]. The need for little conformational change generally allows channels to function at higher rates than transporters. While membrane channels perform their function generally in very high speeds, they also provide impressive specificity and selectivity for certain species. The high selectivity is a vital part of the function of membrane channels, since the intactness of the cell and preservation of the established gradients across the membrane depend on the channels ability to distinguish its specific substrate from the rest [32,33]. Despite the availability of crystal structures for many channels [34–37], in many cases, description of how a channel achieves both a high conductivity and specificity requires dynamical description of both the channel and the transported species [37–42]. The higher rates and relatively less conformational changes required in the functioning of the channels have made them a target for investigation with molecular dynamics simulations from early on [32, 33, 37, 43–47].

1.3 Specific Membrane Transport Proteins Studied

Elucidating the transport mechanisms and substrate selectivity in both membrane transporters and channels requires dynamical description of the transport process in atomistic detail. In this work, we report our molecular dynamics investigation of three membrane proteins (Fig. 1.2): glycerol-3-phosphate transporter [48], which belongs major facilitator superfamily (MFS), the largest superfamily of secondary active transporters, mammalian urea transporter [49], a membrane channel that is responsible for facilitated diffusion of urea, and finally a water channel, aquaporin, for which the highest resolution crystal structure obtained for a membrane transporter is available [50]. In all three cases, we aimed to elucidate the conformational dynamics of the protein at different scales from side-chain to domain motions and determine the basis of substrate selectivity using calculations of

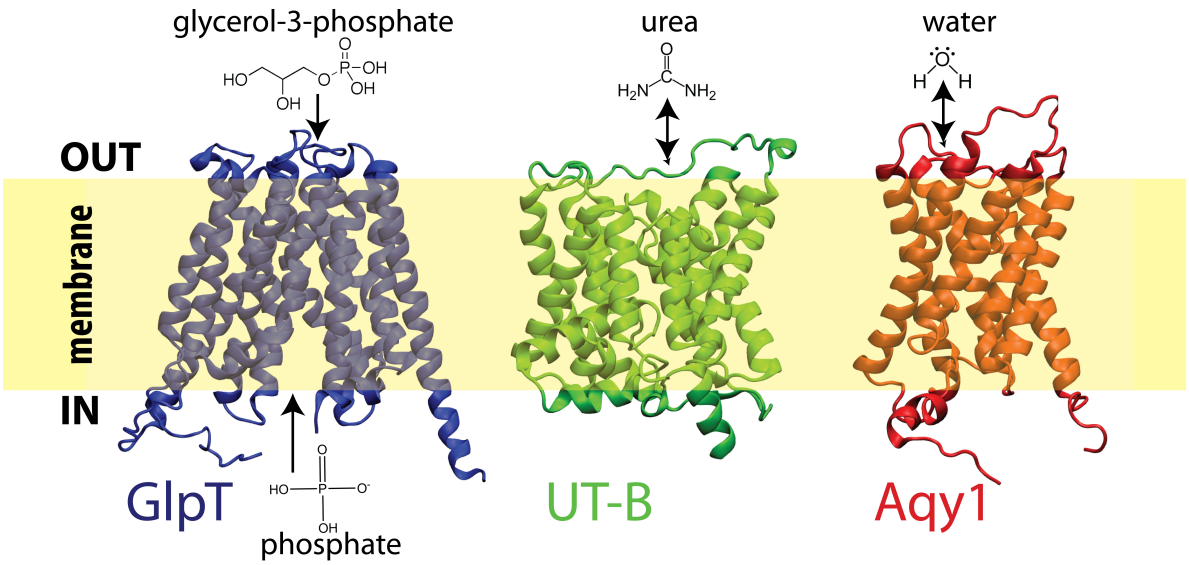

Figure 1.2: All three membrane proteins studied: Glycerol-3-phosphate transporter (GlpT; PDB ID 1pw4 [48]), urea transporter (UT-B; PDB ID 4ezd [49]), and aquaporin (Aqy1; PDB ID 2w2e [50])

the associated energetics.

1.3.1 Glycerol-3-Phosphate Transporter (GlpT)

As the largest superfamily of secondary active transporters and second largest of all transporter families after ATP-binding cassette (ABC) transporters, major facilitator superfamily (MFS) includes ∼25% of all known membrane transport proteins in prokaryotes [51, 52] and consists of 74 families [53]. Found in all kingdoms of life, MFS transporters include several medically and pharmacologically important proteins, e.g., efflux pumps conferring resistance to antibiotics in bacteria and to chemotherapeutics in cancer cells [48, 52, 54–60].

MFS transporters function with three distinct energy coupling mechanisms: uniport, in which the substrate is transported down its electrochemical gradient; symport, in which at least two substrates are transported in the same direction simultaneously fueled by the electrochemical gradient of one substrate; and antiport, in which at least two substrates are transported in opposite directions using electrochemical gradient of one substrate. While individual MFS transporters manifest strict substrate specificity, MFS as a whole handles a diverse set of substrates [48, 52].

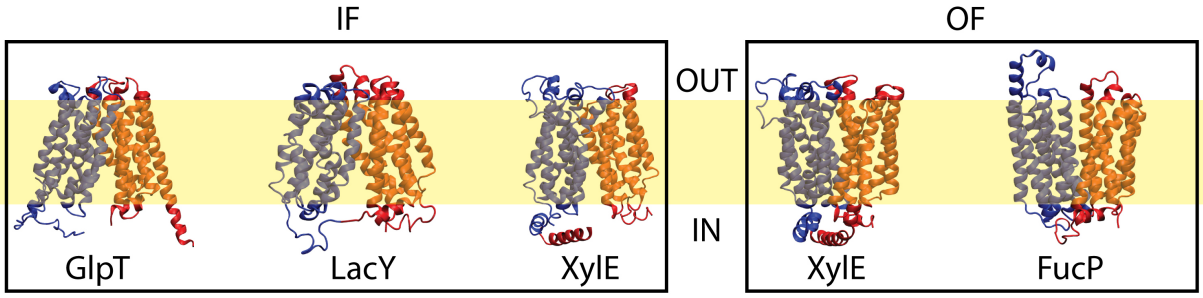

Figure 1.3: Some of the crystal structures available for members of MFS. The position of the membrane is indicated by the semi-transparent yellow band. Glycerol-3-phosphate transporter (GlpT) and lactose permease (LacY) are the first crystallized members of MFS. While both GlpT and LacY were captured in IF state, fucose transporter (FucP) was crystallized much later in the OF state, providing a support for the "rocker-switch"-type conformational change and "alternating-access mechanism". XyLE has recently been crystallized in both IF and OF states, in both substrate-bound and *apo* conformations. All structures have 12 transmembrane helices connected by a loop, which is in some cases structured, and in others were not resolved. The N- and C-terminal halves are colored blue and red respectively.

MFS transporters share a common fold characterized by a single 400-600 amino acid long polypeptide chain folded into a pair of transmembrane bundles, each composed of 6,7, or 12 α-helix bundles connected by a central loop [48, 51, 60] (Fig. 1.3). Recent bioinformatical studies have suggested that triplication and subsequent duplication of a single 2-transmembrane segment hairpin structure evolutionarily gave rise to 12-helix topology of the MFS [53]. The high degree of structural topological type of these permeases. The similarity between the structurally characterized MFS transporters supports a common alternating-access mechanism of IF$\leftrightarrow$OF transition that involves a conserved set of large-scale conformational changes despite their distinct functional/physiological roles. This putative common mechanism of conformational transition involves a *rocker switch*-type conformational change where the two helix bundles rotate with respect to each other, as implied by the structures in different conformational states [61–63]. Following the proposal [64] that MFS transporters share the inverted repeat topology observed in other transporter families [65, 66], a structural model of lactose permease (LacY) from MFS in the OF conformation was obtained by swapping the conformations of inverted-topology repeats found in the protein, which conforms to the alternating access model both theoretically and experimentally [64].

In order to study the transport cycle and determinants of substrate specificity in MFS

transporters, we have concentrated mainly on glycerol-3-phosphate transporter (GlpT; Fig. 1.2), which is, along with lactose permease (LacY), the first MFS transporter crystallized. GlpT mediates the transport of glycerol-3-phosphate (G3P) across the inner membrane of *E. coli*. GlpT belongs to the organophosphate:phosphate antiporter family of major facilitator superfamily (MFS) [48,52,57,60,67–73]. Apart from its role in nutrient uptake, GlpT is also associated with the uptake of the antibiotic fosfomycin, (a G3P-analog), such that bacterial strains with dysfunctional GlpT mutants exhibit fosfomycin resistance [55, 74–77]. Moreover, as one of the first structurally available transporters in MFS, GlpT is highlighted as a structural and functional model for other MFS proteins and has been used as template in homology modeling of several eucaryotic MFS transporters, including its human homologs, hexose facilitator (GLUT1) [73, 78] and glucose-6-phosphate transporter (G6PT), [79], mutations in which cause glycogen storage disease type 1b (GSD-1b) [80,81]. We report our extensive set of equilibrium simulations, in which we identified the binding pathway and mechanism as well as substrate-binding induced conformational changes (Chapter 3), the key amino-acids involved in binding specificity (Chapter 4), the water-conducting states observed in simulations (Chapter 5), and finally, characterization of IF↔OF transition and associated free energies (Chapter 6).

1.3.2 Urea Transporter (UT)

Urea is a small molecule ubiquitously found in all life forms. It is not only a nitrogen source for bacteria, but also a nontoxic carrier of excess nitrogen produced as a result of protein catabolism in mammals [82, 83]. Excretion of urea by kidneys requires little water due to its high solubility and low toxicity. Nevertheless, the osmotic pressure generated by urea in the collecting ducts of the kidneys can potentially draw water from surrounding tissues to urinary space resulting in water loss. To facilitate the process of water reabsorption, urea transporters (UTs) allow urea concentration to rapidly equilibrate between the urinary space and the surrounding tissues [82, 84–86]. Recently, our collaborators, Ming Zhou lab (Columbia University, NY)[†] crystallized both a bacterial [83] and a mammalian [49]

[†]Current position at Baylor College of Medicine, Houston, TX

homolog of urea transporter (Fig. 1.2).

Urea transporters (UTs) are a family of integral membrane proteins that mediate the rapid and passive diffusion of urea down its concentration gradient. In mammals, UTs are expressed in a wide variety of tissues, but their function is best understood in the kidney where they contribute to maintaining the high interstitial urea concentration necessary to limit the rate of water loss [85–87]. During periods of water deprivation, the kidney develops a steep urea gradient from the cortex at 5–8 mM (roughly the concentration present in plasma), to as much as 100-fold higher in the inner medulla [88]. While active transport of ions out of the renal tubules is currently thought to provide the main energetic driving force for creating this gradient, passive transport through UTs also contributes through a countercurrent exchange mechanism that slows the diffusion of urea away from the inner medulla [89]. Additionally, UTs expressed in the inner medullary collecting ducts allow the rapid equilibration of urea between the lumen and the interstitium, preventing water loss driven by the high concentration of urea present in the urine [90]. The importance of UTs in the urinary concentrating mechanism has been verified by extensive knockout studies in mice [91–95], and mutations in UT genes in humans have been linked to variations in blood pressure [96] and the incidence of bladder cancer [97, 98]. Moreover, UT inhibitors, as a new class of diuretics called "urearetics", may serve as potential therapeutics for diseases involving water retention, such as heart failure, cirrhosis, and hypertension, as well as in patients with low anti-diuretic hormone (ADH) secretion [82, 96, 99].

Two genes encode for UTs in mammals: *slc14a1* and *slc14a2*. The *slc14a1* gene contains a single UT domain encoding the protein UT-B, which is expressed in the vasa recta, the nephron's primary blood vessel, as well as in a number of other tissues including erythrocytes, heart, colon, and the brain [100]. In contrast, the *slc14a2* gene, which encodes UT-A, contains two UT domains in tandem, produces a variety of isoforms via alternative splicing, and is regulated by phosphorylation induced by the antidiuretic hormone vasopressin [101–103]. Both UTs facilitate permeation of urea down its concentration gradient.

Despite being called "transporters", the high transport rate of UTs and the available crystal structures suggest a channel-like mechanism [49, 83]. The details of such a mechanism, however, requires investigation of the protein and substrate dynamics in a natural environment. We performed molecular dynamics simulations to investigate urea permeation and associated free energies (Chapter 7) [49]. Besides, in collaboration with Walter F. Boron lab (Case Western University, Cleveland), we calculated gas permeability of UTs and characterized UTs as the third family of gas channels along with aquaporins, water channels, and Rh-associated glycoprotein (RhAG) (ammonia channels).

1.3.3 Aquaporin (Aqy1)

Aquaporins are water transport facilitators found in all kingdoms of life [104]. They are primarily responsible for water homeostasis within living cells, although a subset of aquaporins also facilitates the flow of other small polar molecules, such as glycerol or urea. As with any membrane transport facilitator, aquaporins have evolved to be highly selective for their transported substrate without binding water so strongly that transport is inhibited. In addition to excluding hydroxide (OH^-) and hydronium (H_3O^+) ions, aquaporins must also prevent proton transport via a Grotthuss mechanism [105] in which protons are rapidly exchanged between hydrogen bonded water molecules.

Crystal structures of bacterial [106, 107], archael [108], yeast [109], plasmodium [110], plant [111], mammalian [112–114], and human [115–117] aquaporins have established that these channels contain six transmembrane α-helices and associate as homo-tetramers. A seventh pseudo-transmembrane helix is formed by two loops, which fold as aligned half-helices that insert from opposite sides of the membrane and place the conserved dual asparagine-proline-alanine (NPA) signature motif near the center of the water pore (Fig. 9.1). Transport specificity is defined by the aromatic/arginine selectivity filter (SF) [118, 119], which is located near the extracellular pore entrance and forms the narrowest portion of the channel.

Several models accounting for the ability of aquaporins to impede the passage of protons have emerged from structural arguments [115], molecular dynamics (MD) investiga-

tion of water structure and dynamics [119,120], and computational studies characterizing the energetics associated with explicit transfer of protons across the channel [121, 122]. These studies assert different microscopic mechanisms for excluding protons, including electrostatic repulsion [121,123–125], configurational barriers [119], and desolvation penalties [126], and they consistently report the NPA region, where the macrodipoles of the two half-helices focus a positive electrostatic potential, as the main barrier against proton transport [119–126]. This creates an electrostatic barrier to proton transport [123] and orients the water molecule's dipole moment near the NPA motif, such that the order of oxygen and hydrogen atoms do not support proton exchange via a Grotthuss mechanism [119, 120]. Although intuitively appealing, this picture does not explain why mutations within the NPA motifs that diminish this positive electrostatic barrier facilitate the transport of sodium ions, but not protons [127, 128]; nor is it evident why mutations within the SF can allow the channel to conduct protons [129, 130].

Our collaborators, Richard Neutze lab (University of Gotenburg, Sweden), crystallized the only aquaporin of yeast species, *Pichia pastoris* at 0.88 Å resolution (Fig. 1.2). The high resolution structure reveals hydrogen-bonded network of water and protein side chains. We performed molecular dynamics simulations of this structure to elucidate the dynamics of water transport, which allowed us to put forward a new mechanism for proton-exclusion explaining the effect of the mutations.

Chapter 2

Overview of the General Methodology

2.1 Molecular Dynamics (MD)

The molecular dynamics (MD) is a method largely based on classical and statistical mechanics [131]. Intensive calculations are required in this method, hence the efficiency of MD simulations is heavily dependent on algorithmic developments in mathematics and computer science as well as on advancements in computer hardware [132]. In a classical MD simulation interactions between particles are calculated based from a predefined a potential function, "a force field". 'Empirical force field" parameters are derived from experiments and quantum mechanical calculations, to reproduce chemical and thermodynamic properties of the molecule, but are not able to capture chemical reactions. A force-field typically used in biomolecular simulations contain bonded (bonds, angles, dihedrals) and non-bonded (van der Waals and electrostatic) interaction terms:

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{vdW} + U_{elec}$$

The bonded terms are represented with an harmonic function, while non-bonded interactions usually are usually more complicated and computationally expensive.

$$U_{elec} = k_e \frac{q_i q_j}{r}; \quad V_{vdW} = \varepsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}]$$

The force, $\vec{F}$, acting on each atom is calculated as the negative gradient of the potential energy, U with respect to its coordinates, $\vec{r}$:

$$\vec{F} = -\vec{\nabla} U$$

The Newtonian equations of motion are then integrated [131,133] to propagate positions, $\vec{r}$, and velocity, $\vec{v}$, of each atoms in time, t:

$$\vec{F} = m\vec{a}; \quad \vec{a} = \frac{d\vec{v}}{dt}; \quad \vec{v} = \frac{d\vec{r}}{dt}$$

where m is the mass and $\vec{a}$ is the acceleration, resulting in a trajectory that includes an ensemble of configurations of the system over a period. Several thermodynamic and dynamic properties of the systems can be calculated from the ensemble of conformations obtained from MD simulations.

MD provides dynamical information at high spatial and temporal resolutions and allows testing out various conditions that cannot be tested experimentally. On the other hand, MD is limited by the accessible time scales. Many biological processes of interest take place in longer than millisecond time scales. However, simulations of large biomolecular systems are currently limited to microsecond timescales and most of the time even shorter. The timescale limitation usually results inadequate sampling of the configuration space. Several biased sampling methods have been developed to overcome the problem of limited sampling [134–141]. Besides, the use of simplified potential energy functions do not necessarily capture electronic properties, such as polarization effects. In order to solve this problem, "polarizable" force fields are being developed [142–145].

Despite these limitations, MD has been successfully employed in studying a wide

range of biomolecular systems and phenomena, including membrane transport proteins and their mechanisms [2,32,146]. Furthermore, with continuous algorithmic improvement, availability of faster hardware, better force fields, and enhanced sampling techniques, the gap between simulations and experiments is rapidly closing, as evidenced by recent studies reporting simulations reaching timescales on the order of μs-sub-ms [147–150].

2.2 Free Energy Calculations

The calculation of free energy differences from MD is one of the main challenges in biomolecular simulations. Most of the time equilibrium MD simulations suffer from lack of sufficient sampling due to the system's rugged energy landscape and the shortness of the time scale captured in MD.

2.2.1 Umbrella Sampling (US)

Umbrella sampling is an enhanced sampling method used to sample low probability events [151, 152] and calculate the free energy along a reaction coordinate, namely the potential of mean force (PMF):

$$\Delta F(\zeta) = -k_B T \ln(\rho(\zeta))$$

, where ΔF is the free energy, k_B is the Boltzmann constant, T is the temperature, and $\rho(\zeta)$ is the probability distribution along a reaction coordinate of interest, ζ. In umbrella sampling, the reaction coordinate is divided into several windows. Each window is, then, simulated, independently with a different artificial biasing potential, often in harmonic form added to the system's Hamiltonian:

$$U_i(\zeta) = \frac{1}{2} k_i (\zeta - \zeta_i)^2$$

, where i is the index, k_i is the force constant, and ρ_i is the center of the window. The added bias is designed to flatten the energy barriers, and consequently, improve sampling

in a certain region of the reaction coordinate.

2.2.2 Bias-exchange Umbrella Sampling (BEUS)

Employing conventional umbrella sampling as described above to large-scale transitions is often challenging and may produce unreliable estimates for free energies, especially in large conformational changes. One of the methods to improve convergence of free energy estimations from the umbrella sampling simulations and to ensure the continuity of the phase space sampled is to employ a replica-exchange scheme [153, 154]. This scheme is referred as bias-exchange umbrella sampling (BEUS) (also as window-exchange or replica-exchange umbrella sampling [154–156]). In bias exchange umbrella sampling, the individual windows, each with a different bias, are allowed to exchange potentials based on a Metropolis criterion that preserves the detailed balance. The exchange probability is $P_{ij}^{exchange} = \min(\exp(-\Delta E/k_BT), 1)$, where $\Delta E = (U_i(\mathbf{X}_i^t) - U_i(\mathbf{X}_j^t)) + (U_j(\mathbf{X}_j^t) - U_j(\mathbf{X}_i^t))$, i and j are indices of the windows, $U_i(\mathbf{X})$ is the biasing potential for configuration $\mathbf{X}$ according to umbrella i, and $\mathbf{X}_i^t$ and $\mathbf{X}_j^t$ represent two different configurations whose biases are attempted to be exchanged at time t.

2.2.3 Reconstruction of the Potential of Mean Force (PMF)

To reconstruct the PMF, the biased probability densities calculated from each window can, then, be unbiased and combined using weighted histogram analysis method (WHAM) [157]. In WHAM, the following equations are solved iteratively to self-consistency:

$$\begin{cases} \rho(\zeta) = \dfrac{\sum_{i=1}^{N_{win}} n_i(\zeta)}{\sum_{i=1}^{N_{win}} N_i e^{(f_i - U_i(\zeta))/k_BT}}, \\ f_i = -k_BT \ln(\sum_{bins} \rho(\zeta) e^{-U_i(\zeta)/k_BT}), \end{cases}$$

where N_{win} is the total number of windows, $n_i(\zeta)$ is the histogram counts, and $U_i(\zeta)$ is the biasing potential in bin associated with *zeta* and window i, $\rho(\zeta)$ is the estimate of the unbiased probability distribution.

One of the limits of the regular WHAM is that PMF can only be reconstructed as a function of the reaction coordinate biased in the umbrella sampling scheme. A generalization of both WHAM and Bennett acceptance ratio (BAR) [158], called Multistate Bennett Acceptance Ratio or generalized WHAM (GWHAM) [140], on the other hand, allows assignment of weights to individual configurations $\mathbf{X}_i^t$, thus, reconstruction of PMF as a function of any reaction coordinate [159, 160] The weight of each configuration $\mathbf{X}_i^t$, p_i^t, is determined by iteratively solving the following equations to self-consistency [159]:

$$\begin{cases} 1/p_i^t = & \sum_j N_j f_j \exp(-\beta U_j(\mathbf{X}_i^t)), \\ 1/f_j = & \sum_i \sum_{t=1}^{N_i} p_i^t \exp(-\beta U_j(\mathbf{X}_i^t)), \end{cases}$$

in which $U_i(\mathbf{X})$ is the biasing potential for configuration $\mathbf{X}$ according to umbrella i. PMF in terms of any reaction coordinate can, then, be constructed by measuring any property of the configuration $\mathbf{X}_i^t$ and reweighting it with p_i^t. GWHAM has been the preferred method of analysis for BEUS, and regular WHAM for the conventional umbrella sampling

2.3 System Preparation of Membrane Protein Systems

In a typical simulation of membrane proteins that we present here, experimentally solved atomistic protein structure is placed in water, lipids (membrane), and ions. In order to mimic experimental conditions, MD simulations are often carried out under NPT (constant-temperature (300–310 K) and constant-pressure (1 atm), constant number of particles) Force field parameters are available for standard molecular systems, such as protein residues, nucleotides, and lipid molecules [161, 162]. Missing parameters for ligands are included by adopting similar parameters from the available force fields. The simulations involve a brief period of initial membrane equilibration (typically 1–5 ns), wherein the lipid tails are allowed to equilibrate while the lipid head groups and the protein are constrained to their initial positions. This is then followed by an unconstrained equilibration of the lipids and the protein for simulation times ranging tens to hundreds of

nanoseconds depending on the system and the problem. MD simulations, presented in this work, were performed using supercomputers from the Extreme Science and Engineering Discovery Environment (XSEDE), and the campus clusters of University of Illinois using the program NAMD [163]. The simulation trajectories were analyzed with the program VMD [164].

Chapter 3

Substrate Binding and Initial Conformational Response of GlpT*

3.1 Introduction

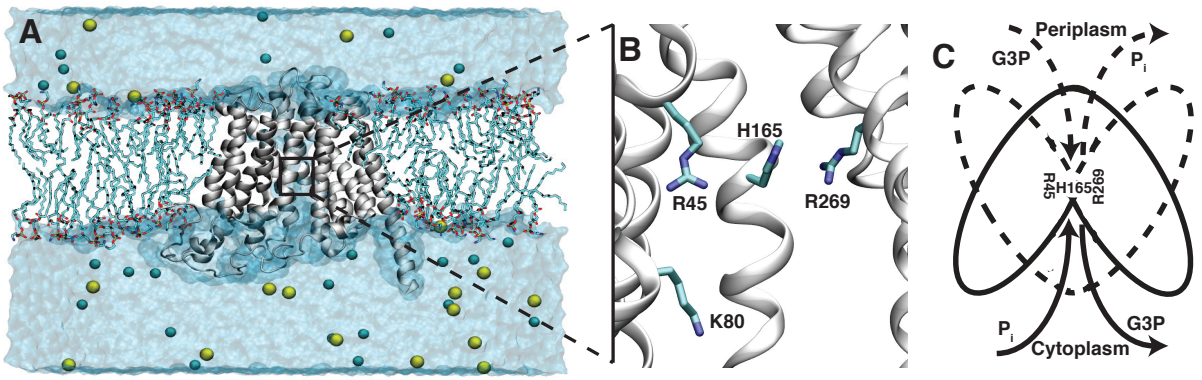

Figure 3.1: Structure, simulation system, and the schematic mechanism for GlpT. (A) Simulation system constructed using the crystal structure of GlpT (cartoon representation), embedded in POPE membrane (partially shown for clarity), water (transparent) and the ions (spheres). (B) The putative substrate-binding site of GlpT suggested based on the crystal structure of the *apo* protein. Key residues, namely, R45, K80 and H165 on the N-terminal half and R269 on the C-terminal half are shown. (C) Rocker-switch / alternating-access model. Solid lines and dashed lines indicate the crystallized cytoplasmic-open (IF) state and the hypothetical periplasmic-open state of GlpT, respectively. P_i binding from the cytoplasmic side results in conversion to the periplasmic-open state, where G3P replaces P_i. At any given time the single binding site is accessible only from one side.

*This chapter appeared as a research article in Biophysical Journal [165]. Giray Enkavi and Emad Tajkhorshid, "Simulation of spontaneous substrate binding revealing the binding pathway and mechanism and initial conformational response of GlpT". *Biochemistry*, 49:1105–1114, **2010**.

Structurally, GlpT is organized into two transmembrane-helix bundles, the N- and C-terminal halves, each composed of 6 α-helices. The bundles exhibit a pseudo-twofold symmetry with weak sequence homology (Fig. 3.1A) [48, 55, 166]. The cytoplasmic-open (IF) structure captured in the crystal structure [48] revealed a lumen formed between the two halves, which is closed on the periplasmic side and open on the cytoplasmic side. It was proposed that this lumen provides the pathway for substrate translocation, and the apex of the lumen was implicated as the putative substrate-binding site (Fig. 3.1A and B). The putative site includes two pseudo-symmetrically positioned, highly conserved arginines (R45 and R269), on the N- and C-terminal halves, respectively, as well as a histidine (H165) located between them (Fig. 3.1B The proposed binding site and its constituting side chains also shows a high degree of sequence similarity to hexose-6-phosphate transporter (UhpT), a more extensively studied homolog of GlpT [48]. Indeed, UhpT residues corresponding to R45 and R269 in GlpT are functionally indispensable, whereas other arginines in UhpT can be mutated to lysines without significant loss of function [48, 55, 57, 166–169]. The functional significance of the residues in the putative binding site has also been confirmed by mutagenesis experiments directly performed on GlpT [170].

An "alternating-access" / "rocker-switch mechanism" has been proposed for the transport cycle of GlpT, which appears to function as a monomer under physiological conditions [48, 167]. According to the proposed model, under physiological conditions P_i binding from the cytoplasmic side results in the closure of the cytoplasmic vestibule of the lumen and opening of the periplasmic side (formation of the periplasmic-open state). Replacement of P_i by G3P in the periplasmic-open state induces the returning of GlpT to its initial state, through a reverse set of protein conformational changes (Fig. 3.1C). These conformational changes are suggested to involve substrate-induced weakening of interactions between the N- and C-terminal halves on the initially closed side along with the formation of new interactions on the opposite side [48, 55, 57]. Kinetic studies have shown that the interconversion of the states is rate limiting and temperature-dependent, indicating the involvement of large protein conformational changes, whereas substrate

binding is found to be temperature-independent and rapid. Accordingly, substrate binding is proposed to accelerate the interconversion of the states by lowering the activation energy [60]. Starting from the crystal structure, 10 degrees of rigid-body rotation of each half was shown to be sufficient to open the periplasmic side and close the cytoplasmic side [48,52,55,57]. Also, a computational study suggested that 9–10 degrees of rigid-body rotation generates a conformational state with maximal electrostatic interaction between the N- and C-terminal halves [171].

3.2 Methods

3.2.1 System Preparation

The crystal structure of *apo* GlpT [48] as deposited in the RCSB Protein Data Bank [172], PDB ID 1pw4, was adopted as the initial structure for all the simulations. The protonation state of the titratable side chains was determined using the MolProbity website [173,174]. Accordingly, the binding site histidine, H165, was modeled in its uncharged form (with N_δ protonated. The mutated residues in the crystal structure were reverted back to the wild-type ones, and the missing side chains were added using the PSFGEN plugin of VMD [164]. The missing inter-domain loop, which connects the N- and C-terminal halves, was modeled as an unstructured chain and partially relaxed through a short, *in vacuo* simulation (100 ps) with the rest of the protein fixed. Cavity water molecules were then added using DOWSER [175], and the protein was solvated using the program Solvate [176]. GlpT was reoriented for correct membrane insertion based on the OPM (Orientations of Proteins in Membranes) database [177] and water molecules in the potential lipid protein interface were deleted. Solvated GlpT was then inserted into a patch of POPE (1-palmytoil-2-oleoyl-*sn*-glycero-3-phosphatidylethanolamine) membrane (115×115 Å^2) generated using the MEMBRANE BUILDER plugin of VMD [164] with the membrane normal along the z-axis such that the protein is at least $\sim$20 Å away from the boundaries of the water box in all directions. The lipid molecules overlapping the protein were deleted. Additional solvent was then added and the system was neutralized with 100 mM NaCl using the

SOLVATE and AUTOIONIZE plugins of VMD [164]. The final dimensions of the system are $115 \times 115 \times 110\,\text{Å}^3$ including ∼125,000 124,719 atoms. While fixing all the atoms except the lipid tails, the system was first minimized for 5000 steps and simulated for 1 ns under NVT (constant volume and temperature) conditions in order to allow the lipid acyl tails to adopt a less ordered conformation (referred to as "melting" the lipid tails). Then, the heavy atoms of the α-helices were constrained using harmonic potentials ($\text{k} = 7.2\,\text{kcal/mol/Å}^2$), and the system was minimized for 5000 steps and simulated at 1 atm pressure for 1.5 ns under NPT (constant pressure and temperature) conditions. Keeping only the backbone atoms of the α-helices constrained in the next step, the system was further minimized for 5000 steps and equilibrated at constant normal pressure (only along the z direction, constant area; $\text{NP}_\text{n}\text{T}$) for 5 ns to allow the membrane, water and ions to adapt to the crystal structure of GlpT. After releasing all the constraints, the system was further equilibrated for 5 ns using the same conditions. The resulting model was then used as the starting configuration of all the other simulations except System 9 (see below and Table 3.1).

3.2.2 Simulation Systems

The simulated systems along with a short description of the substrate behavior are summarized in Table 3.1. The equilibrated membrane-embedded model of GlpT (described above) was simulated for additional 50 ns using free dynamics in two independent simulations to serve as the control (Systems 1 and 2). In order to investigate the mechanism and pathway of substrate binding for different titration states of the substrates using unbiased simulations, spontaneous binding simulations were designed (Systems 3–18). Although under physiological conditions the net effect of the transport cycle in GlpT is the uptake of G3P at the cost of the export of P_i (Fig. 3.1C), GlpT is known to function in either direction, and the directionality of transport is merely determined by the concentration gradient of the substrates [48, 178]. Here, we will study the binding of both P_i and G3P to the cytoplasmic side of the transporter, thus, investigating the behavior of structurally different substrates, which allow us to better probe the common binding site that is pro-

Table 3.1: Simulated systems.

System	Substrate	t (ns)	initial [a] distance (Å)	Neutralized residue[c]	Behavior of the substrate
1	–	**50.0**	–	–	–
2	–	**50.0**	–	–	–
3	$\mathbf{P_i^-}$	**50.0**	**14.82**	–	binds spontaneously
4	P_i^-	∼1.5	19.32	–	binds spontaneously
5	$\mathbf{P_i^{2-}}$	**50.0**	**14.82**	–	binds spontaneously
6	P_i^{3-}	∼2.2	14.82	–	diffuses out
7	P_i^{3-}	∼2.2	14.82	–	diffuses out
8	P_i^{3-}	∼8.3	10.95	–	diffuses out
9	P_i^{3-}	10.0	4.34	–	remains bound
10	$\mathbf{G3P^{2-}}$	**50.0**	**14.82**	–	binds spontaneously
11	$G3P^{2-}$ [b]	∼4.9	13.17	–	binds spontaneously
12	$G3P^{2-}$ [b]	∼3.2	18.49	–	diffuses out
13	$\mathbf{G3P^-}$	**50.0**	**10.89**	–	binds spontaneously
14	$G3P^-$	∼10.0	14.82	–	does not translocate
15	P_i^-	∼3.6	14.82	R45	diffuses out
16	P_i^-	∼1.5	14.82	K80	diffuses out
17	P_i^-	∼5.0	14.82	R269	does not translocate
18	P_i^{2-}	∼0.9	14.82	K80	diffuses out

[a] The initial position of the substrate, specified as the distance between the P atom of the substrate and the C_ζ atom of R45.

[b] In contrast to the other binding simulations where, the phosphate moiety faces the apex of the lumen in its starting configuration, the substrate was placed with the phosphate moiety of G3P facing the cytoplasm, i.e., away from the protein.

[c] Residues were neutralized as described in Section 3.2.2.

posed for GlpT. The substrate-binding simulations were prepared by placing a substrate in the cytoplasmic vestibule of the lumen within an initial distance of 10–15 Å from the putative binding site. The distance, which is defined as that between the P atom of the substrate and C_ζ atom of R45 is specified in Table 3.1 for all the simulation systems. For each substrate (P_i^-, P_i^{2-}, $G3P^-$, and $G3P^{2-}$) 5000 steps of minimization and 50 ns of simulation were then performed (Systems 3, 5, 10, and 13, respectively).

Several additional, shorter simulations were also performed, in order to examine the reproducibility of the phenomena observed in longer simulations, and to study the effect of various modifications and different initial positions and orientations of the substrates (Systems 4, 7, 8, 11, 12, and 14). G3P was placed perpendicular to the plane of the membrane, with the phosphate moiety either facing the apex of the lumen (System 10),

or away from it (Systems 11 and 12). Although physiologically irrelevant, the binding of P_i^{3-} was also investigated using a similar approach (Systems 6, 7, and 8). In order to determine at which level P_i^{3-} is discriminated, in System 9, P_i^{3-} was created in the binding site by removing two protons from an already bound P_i^- (the structure at t = 20 ns taken from System 3).

Four additional simulations were designed to test the electrostatic contribution of key residues to substrate binding, by neutralizing the respective residues starting from System 3 (for Systems 15–17) and System 4 (for System 18). R45 and R269 were neutralized by deprotonation at N_ϵ (thus, leaving the H-bond formation capacity of the terminal NH_2 groups of the guanidinium group unperturbed) and K80 at N_ζ, and adjustment of the atomic charges. Systems 3–18 were each minimized for 5000 steps and then simulated for various times, as listed in Table 3.1. The electroneutrality of the system was preserved by deleting appropriate numbers of Cl^- ions sufficiently far away from the protein, when needed. In the rest of the text, *apo* refers to Systems 1 and 2, and the substrate-binding simulations, P_i^-, P_i^{2-}, $G3P^-$, and $G3P^{2-}$, refer to Systems 3, 5, 10, and 13, respectively (boldfaced in Table 3.1).

3.2.3 Simulation Protocols

The simulations were performed with NAMD 2.6 [163] using the CHARMM27 force field with ϕ/ψ cross term map (CMAP) corrections [161]. All the production simulations were carried out using a 1-fs time-step, at a constant temperature of 310 K, and a constant pressure of 1 atm maintained only along the z direction, i.e., normal to the membrane (NP_nT). Water molecules were modeled as TIP3P [179]. The force field parameters for different protonation states of the substrates (P_i^-, P_i^{2-}, P_i^{3-}, $G3P^-$, and $G3P^{2-}$) were adopted from similar molecules ($Me{-}PO_4^-$, $Me{-}PO_4^{2-}$, and glycerol) in the CHARMM force field. Constant pressure was maintained by the Nosé-Hoover Langevin piston method [180,181], and constant temperature by Langevin dynamics with a damping coefficient of $1\,ps^{-1}$ for non-hydrogen atoms. Short-range interaction cutoff was set to 12 Å. Long-range electrostatic interactions were computed using the particle mesh Ewald (PME) method [182]

with a grid density of at least 1 $Å^{-3}$. Bonded, non-bonded, and PME calculations were performed at 1, 2, and 4 time steps, respectively.

3.3 Results and Discussion

3.3.1 Pathway and Mechanism of Substrate Binding

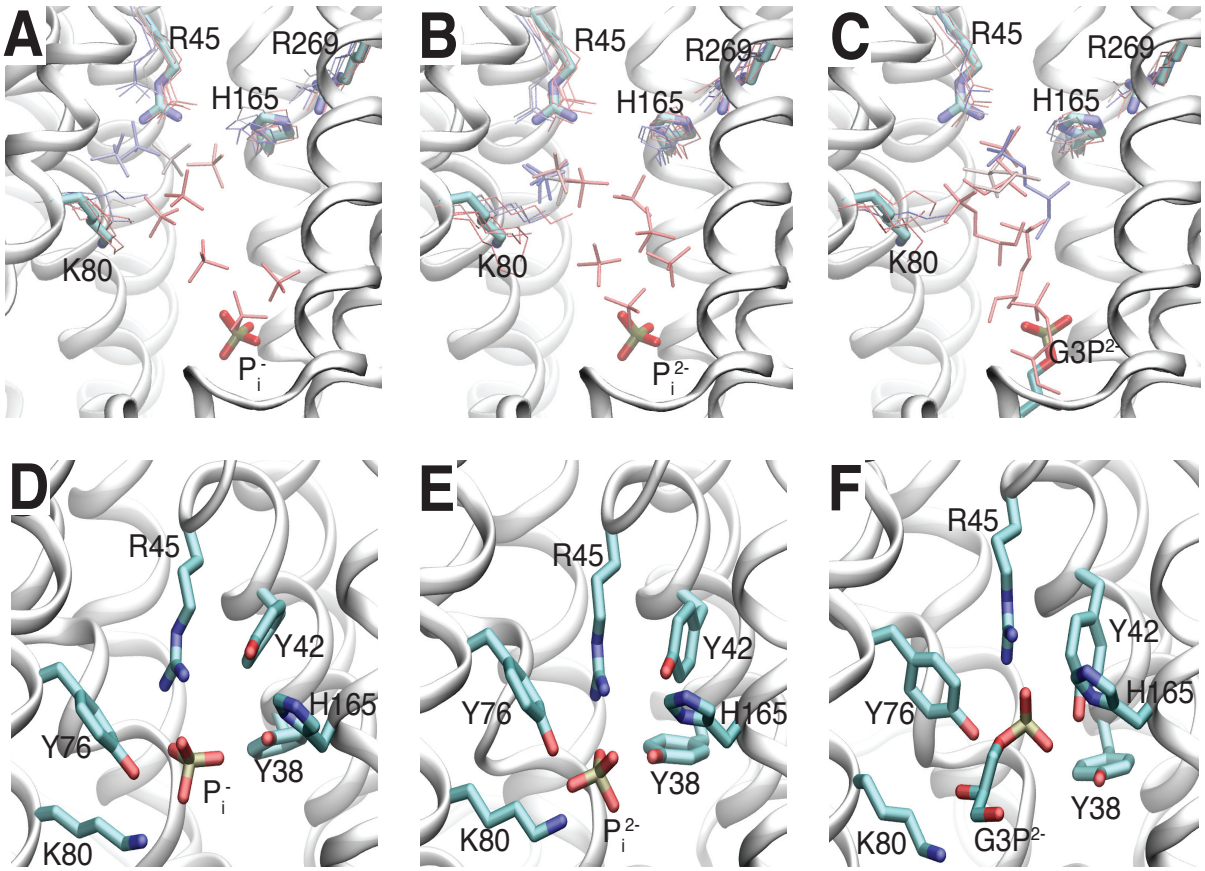

Figure 3.2: Spontaneous binding of P_i and G3P and their bound states. (A–C) The trajectories of the substrates, namely P_i^- (A), P_i^{2-} (B), and $G3P^{2-}$ (C), in the substrate-binding simulations are shown against a static representation of the equilibrated GlpT (t = 0 ns of System 1). The time evolution is represented by color change from red to gray to blue (t=0-50 ns). The initial position of the substrate is shown in a thicker stick representation. (D–F) The substrates are shown in their stably bound configurations for P_i^- (D), P_i^{2-} (E), and $G3P^{2-}$ (F), along with the key residues interacting with them. The view in D–F is rotated from that shown in A–C by about 90 ° around the z-axis, in order to optimally display the binding site residues.

In order to characterize the substrate binding pathway and mechanism, and to investigate substrate recognition, four independent, extended equilibrium MD simulations

were performed in the presence of either substrate (P_i and G3P) at their physiologically relevant protonation states (P_i^-, P_i^{2-}, $G3P^-$, and $G3P^{2-}$) along with several shorter simulations to examine the reproducibility of the observed phenomena (Table 3.1). In order to avoid spending simulation time unnecessarily on the diffusion of the substrate in the solution outside the protein, the substrate was initially placed at the cytoplasmic mouth of the lumen, however at a large enough distance from the putative binding site (see Table 3.1) that allowed the simulation to focus on its diffusion within the protein's lumen and to describe the role of key lining residues in its recruitment and translocation. Each system was then simulated for 50 ns (Systems 3, 5, 10, and 13). As control, GlpT was also simulated in the *apo* state for 50 ns (Systems 1 and 2).

These simulations resulted in rapid, spontaneous substrate translocation inside the lumen toward the apex revealing a common binding pathway and mechanism for different substrates. Several snapshots depicting the trajectories of the substrates within the protein lumen are shown in Fig. 3.2A–C; in Fig. 3.3 and Fig. 3.4 the coordinates of the phosphate groups of the substrate and the time series of the distance between the substrate and key residues are shown. The interaction between the charged residues of GlpT and the phosphate moiety of the substrates appears to be the main driving force steering the substrates toward a common binding site (Fig. 3.5). The first contact with the protein is established through conserved K80 (Fig. 3.5), which seems to act as a "hook" recruiting and escorting the substrate from the mouth of the lumen deep toward its apex. After transient coordination with K80, and the nearby Y76, the phosphate moiety rapidly associates with R45 in the "binding site". The formation of initial hydrogen bonds with R45 takes as little as $\sim$0.5 ns from the beginning of the simulation. In all cases, substrate binding is accompanied by a large increase in interaction energy between the protein and substrate specially for divalent species (Fig. 3.5). Decomposition of the energies confirms that electrostatic interactions as the main constituent of these changes (Fig. 3.7).

In the bound state, the phosphate moiety is stabilized by hydrogen bonds with the guanidinium group of R45, as well as those with the hydroxyl groups of several surrounding tyrosines (Y38, Y42, Y76) (Fig. 3.6). These hydrogen bonds are all maintained throughout

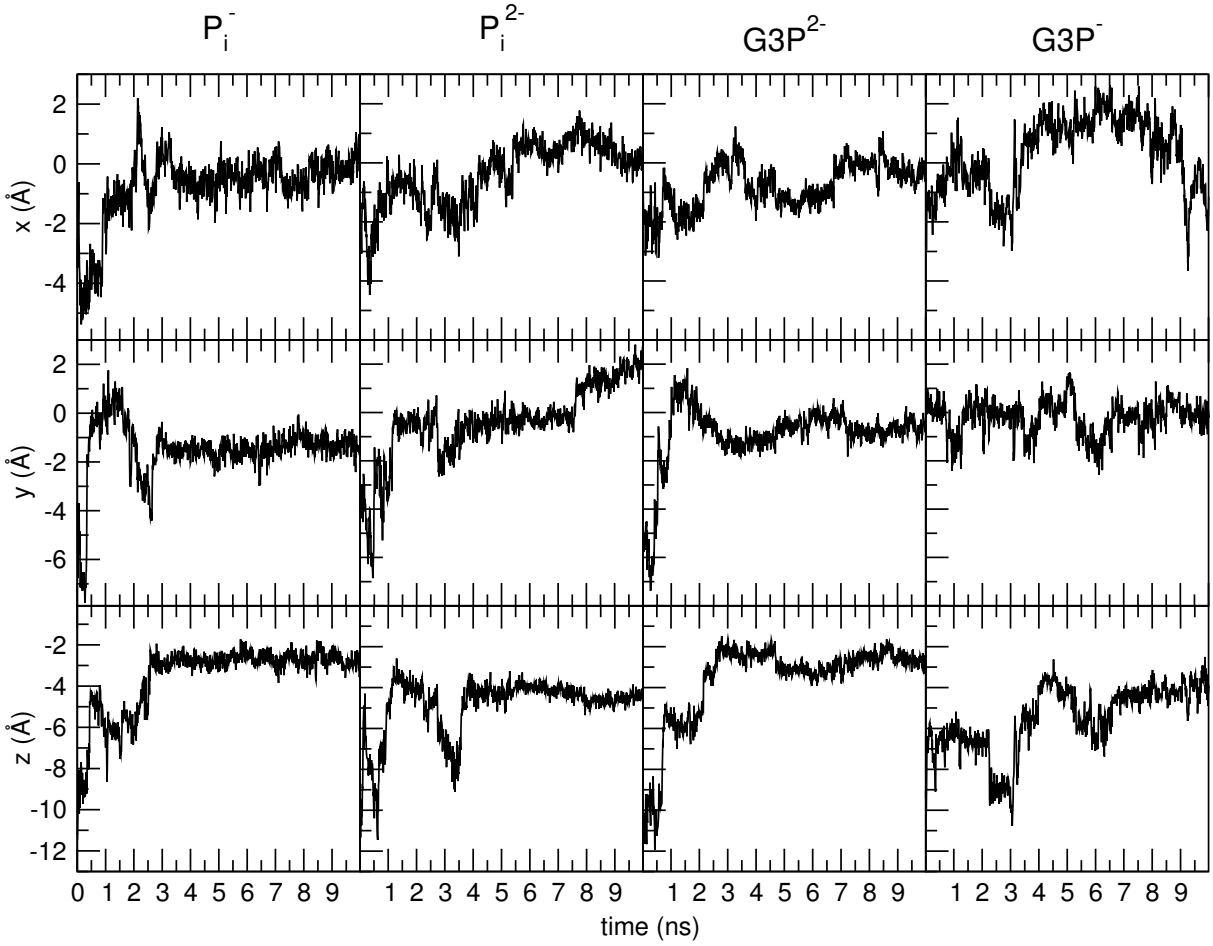

Figure 3.3: x, y, z displacements of the substrates for the first 10 ns of the simulations

the simulations. In P_i^--, P_i^{2-}-, and $G3P^{2-}$-binding simulations, there is almost at all times, at least one hydrogen bond (N–O distance of 2.5 Å or shorter) between R45 and the phosphate moiety after their initial association. $G3P^-$-binding simulation, however, appears to result in a less stable bound state compared to the other substrates, as it exhibits frequent dissociation-reassociation transitions with R45 and larger fluctuations inside the lumen (Fig. 3.4). In fact, $G3P^-$ needed to be initially placed ∼4 Å closer to the apex (in close contact with K80) in order to initiate its initial association and translocation toward the binding site within a comparable time scale to the other substrate-binding simulations (Table 3.1). This might indicate that either $G3P^-$ binding is slower than the other substrates, or $G3P^-$ represents a disfavored protonation state of G3P for transport in GlpT (at least during initial binding). The change in interaction energy for $G3P^-$ as

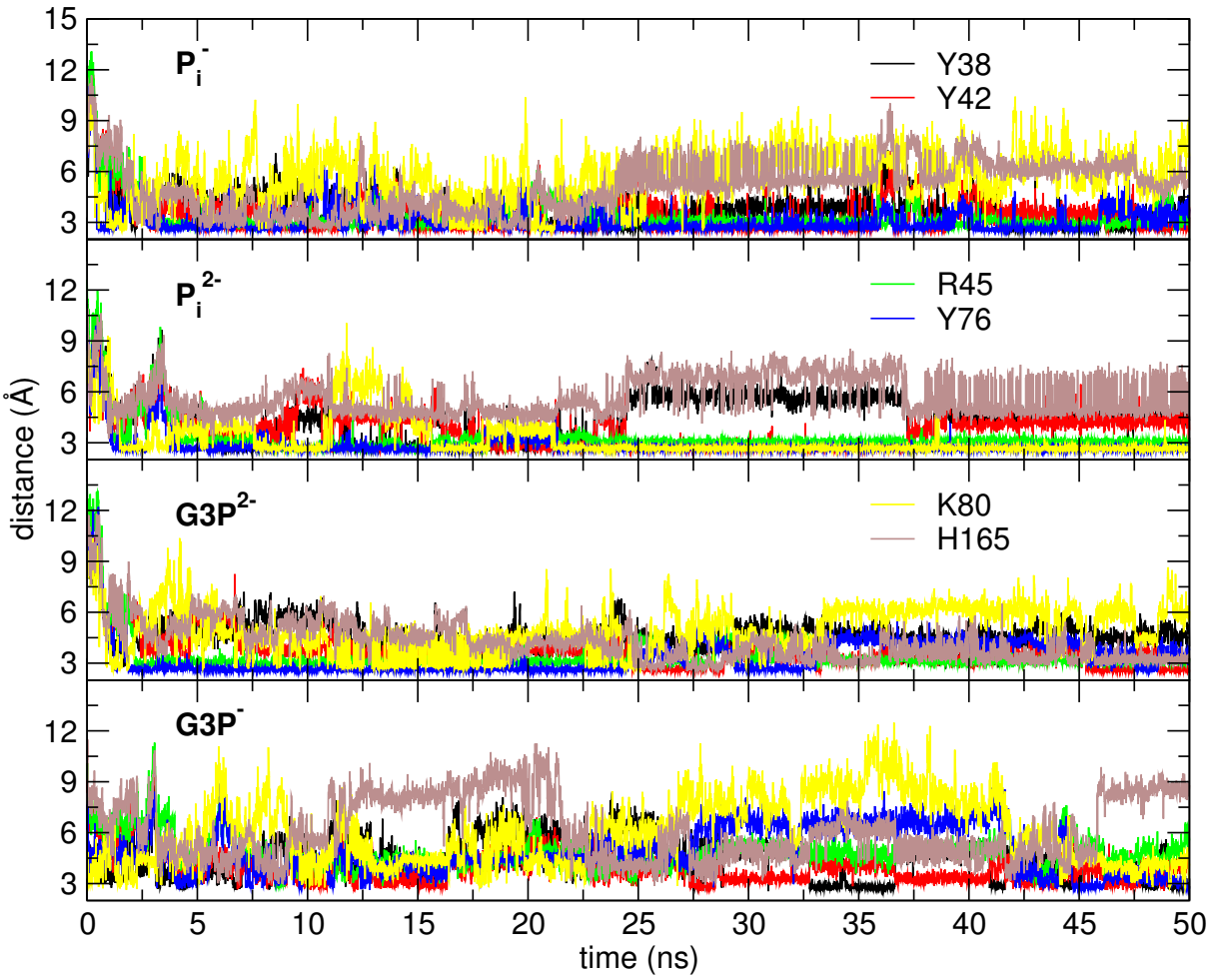

Figure 3.4: Distances of the substrates from the key binding site residues

it moves towards the apex is not only lower than the other substrates but is also flatter (Fig. 3.5), allowing it to sample a larger space in the binding site, instead of tightly binding. Despite the less stable binding of $G3P^-$, many of its binding characteristics, i.e., the order of interaction with the key residues and the type of interactions with the binding site, are similar to those observed for the other substrates.

Several simulations (Systems 15–18) showed that the binding of the substrate is impaired by the neutralization of any of the three key basic residues of the lumen, i.e., R45, K80, and R269 (Table 3.1). Neutralization of either R45 or K80 resulted in P_i^- diffusing out of the lumen, while in the case of R269 neutralization, P_i^- failed to move toward and bind the binding site within 5 ns (stayed in its initial position, ∼15 Å away from R45).

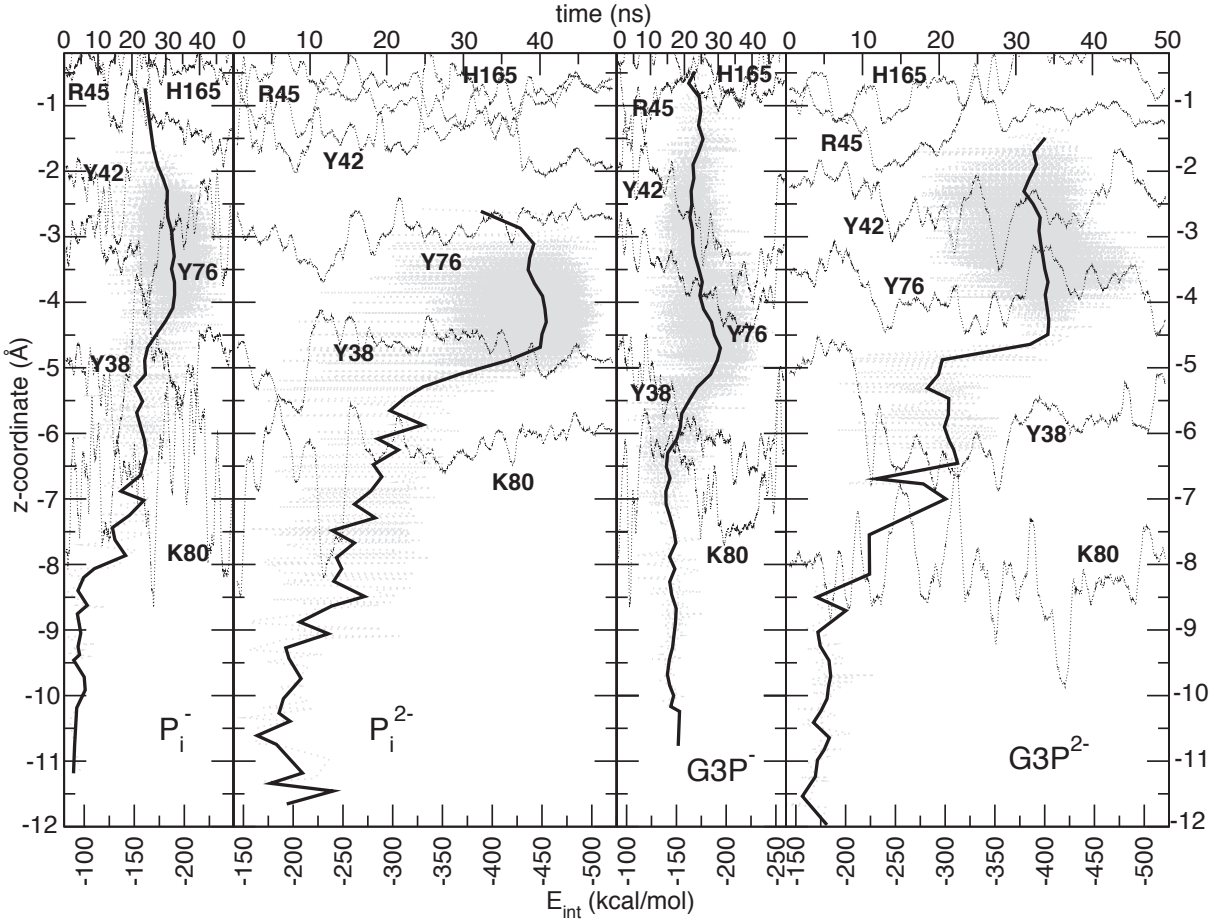

Figure 3.5: Total interaction energies of the substrates with the protein with respect to their z-coordinate calculated from the trajectories. The z-coordinates and the corresponding energies were averaged at every 0.2 Å displacement. The time vs z-coordinate of the geometrical center of N_ϵ of R45, N_δ and N_ϵ of H165, N_ζ of K80, and hydroxyl O atoms of Y38, Y42, and Y76 are also shown (dashed lines, running averages over 1 ns).

Inhibition of spontaneous binding was also observed for P_i^{2-} when K80 was neutralized (System 18). These results clearly indicate that the electrostatic effects of these residues constitute the main driving force in the binding of the substrate to GlpT (Fig. 3.7).

In order to further investigate the role of electrostatics in steering the substrates toward the binding site, simulations in the presence of P_i^{3-} were also performed, noting that P_i^{3-} represents a physiologically irrelevant protonation state of P_i, and, therefore, it has not been viewed as a substrate for GlpT [183,184]. Surprisingly, in all the simulations, P_i^{3-} diffused out of the lumen despite its expected higher electrostatic attraction toward the apex of GlpT. This behavior is most probably due to the strong solvation energy of

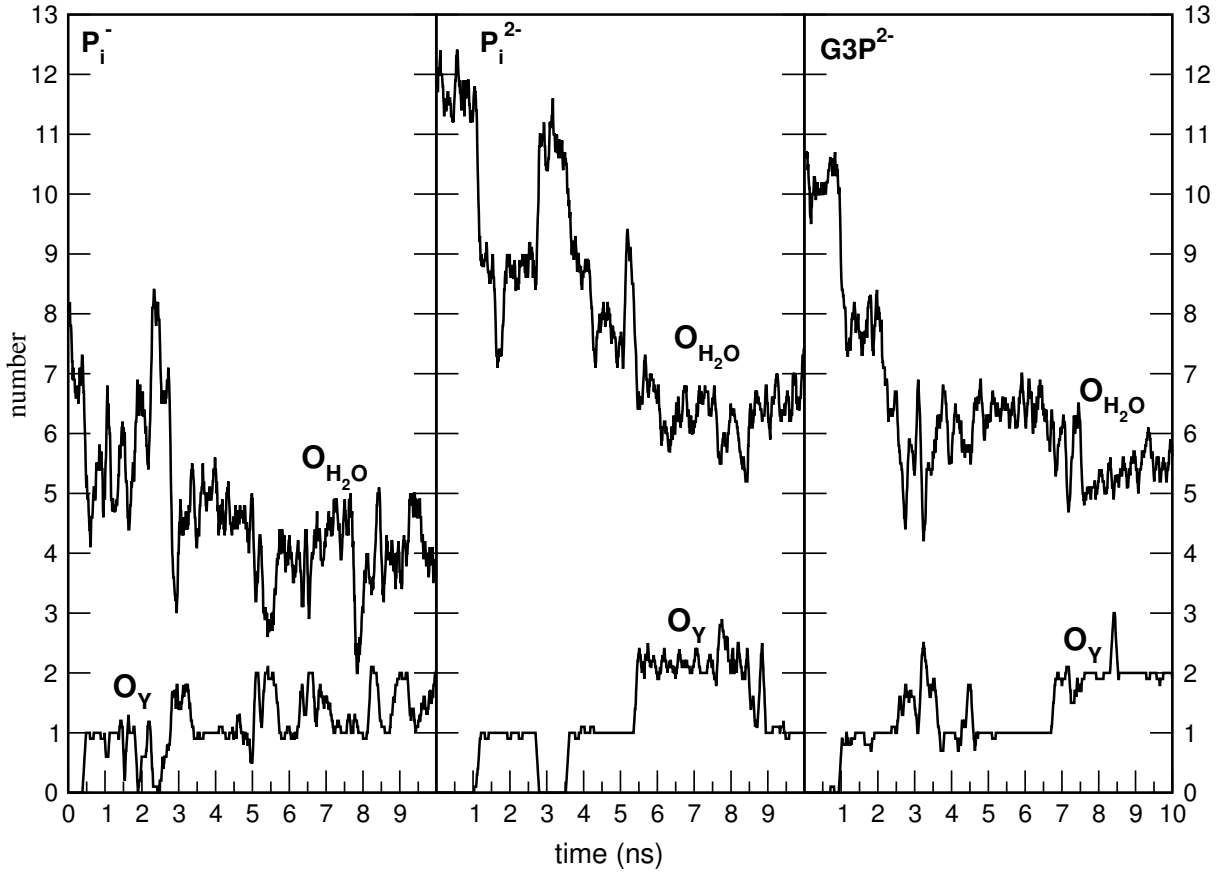

Figure 3.6: Partial dehydration of the substrate in the tyrosine cage. Time series of the number of water oxygen and tyrosine hydroxyl oxygens (Y38, Y42, and Y76) within 3 Å of the phosphate oxygens of P_i^-, P_i^{2-}, $G3P^{2-}$, respectively (Systems 3, 5, and 10). Running averages are calculated over 100 ps and the data is plotted for the first 10 ns of the simulations during which association of the substrate with the binding site takes place.

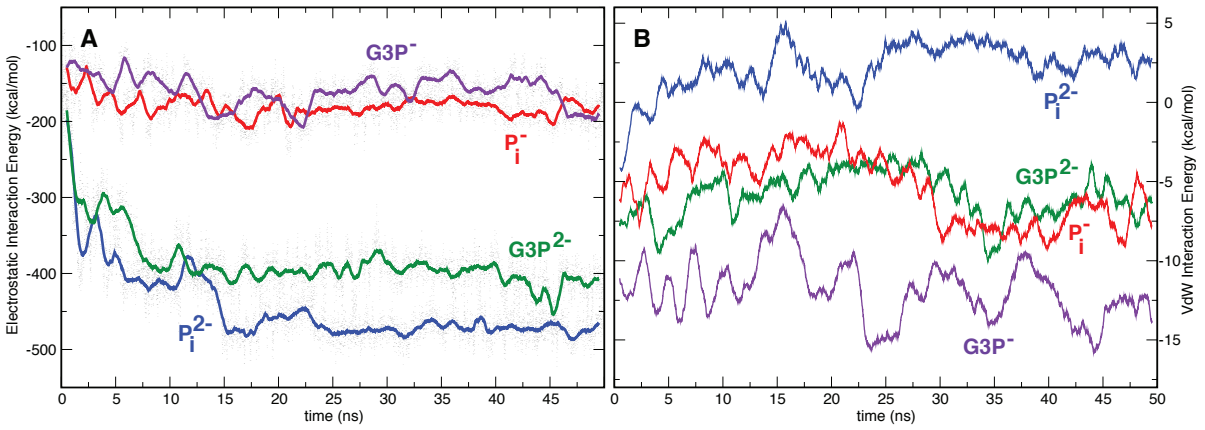

Figure 3.7: Interaction Energies (A) Electrostatic interaction energies between the substrates and the protein. (B) VdW interaction energies between the substrates and the protein.

P_i^{3-} in water, which cannot be overcome by the interaction with the protein. However, deprotonating an already bound P_i^- to P_i^{3-} did not result in unbinding within 10 ns (System 9). Therefore, discrimination against P_i^{3-} is likely enforced during the initial penetration of the substrate into the lumen. Stable binding of P_i^{3-} upon deprotonation of a bound P_i^- is also consistent with the notion that the protonation state of P_i might be modulated after its initial binding.

3.3.2 Details of Interactions in the Substrate Binding Site

The translocation pathway of the substrate, the time scale of initial association and binding, and the final bound states exhibit a high degree of similarity in all substrate-binding simulations (Fig. 3.2D–F, Table 3.4; Fig. 3.3 and Fig. 3.4). The description of the bound state presented in this section is based on P_i^-, P_i^{2-}, and $G3P^{2-}$ simulations in which stable binding was observed (Table 3.2). I note that G3P show two main binding modes dues to its glycerol moiety and will discuss that in the next section In this section, I will mainly focus on the common features of binding, namely those mediated by the interaction between the phosphate moiety and the protein in different substrates. As discussed below in detail, substrate–protein interactions appear to be dominated by electrostatic interactions between the protein and this moiety.

The contribution of the phosphate moiety to binding involves an almost invariant, strong interaction with the guanidinium group of R45, which seems to act like a "fork" holding the phosphate moiety tightly in the binding site during the simulations (Fig. 3.2D–F). The interaction of the phosphate moiety with GlpT in the bound state also involves hydrogen bonding with K80, H165, Y38, Y42, and Y76, with slight variations in different simulations (Table 3.2, Fig. 3.4). K80 seems to frequently lose its direct contact with the phosphate moiety shortly after the formation of the bound state ($t \simeq 2$ ns) (Fig. 3.4). This might suggest that K80 plays its main role in steering the substrate into the binding site and its initial coordination, but becomes less critical once a more stable interaction with R45 has been established. The loss of direct contact with K80 is accompanied by the hydrogen bonding of the substrate with H165. Due to spatial separation between their

Table 3.2: Final distances of the substrate from key residues in the final bound state.[a]

	Y38	Y42	R45	Y76	K80	H165
P_i^-	3.08 ± 0.70	3.37 ± 0.59	3.06 ± 0.38	3.07 ± 0.61	6.19 ± 1.18	6.09 ± 0.65
P_i^{2-}	4.28 ± 0.45	4.51 ± 0.24	3.08 ± 0.12	2.63 ± 0.10	2.67 ± 0.10	5.30 ± 0.88
$G3P^-$	3.84 ± 0.85	3.59 ± 0.62	4.82 ± 0.85	4.17 ± 1.60	5.15 ± 1.53	6.41 ± 1.91
$G3P^{2-}$	4.62 ± 0.51	3.18 ± 0.51	3.10 ± 0.16	3.82 ± 0.61	5.71 ± 0.98	3.43 ± 0.51

[a] The distances are given between the closest nucleophillic atoms on the residues and the phosphate group of the substrate, i.e., hydroxyl oxygens in tyrosine residues, N_ζ in K80, and N_δ in H165. In the case of R45 distance represents the average of distances of the closest phosphate oxygens to N_ϵ atoms. All the means and standard deviations are calculated for the last 10 ns of the simulations. For distance vs. time plots, see Fig. 3.4.

side chains, H165 and K80 do not seem to be able to coordinate the phosphate moiety simultaneously, i.e., at any given time only one of these side chains is directly interacting with the substrate.

The hydroxyl groups of the three conserved tyrosine residues (Y38, Y42, and Y76), which reside below the guanidinium group of R45, are optimally positioned for hydrogen bonding with the phosphate moiety (Fig. 3.2D–F). They form a "cage"-like binding pocket, in which R45 keeps the substrate. Besides, the "tyrosine cage" might aid in desolvation of the substrate and preventing it from diffusing out of the binding pocket. Analysis of the trajectories shows that indeed the formation of new hydrogen bonds between the tyrosine cage and the substrate coincides the partial dehydration of the latter (Fig. 3.6). The role of these tyrosines were previously thought to be limited to maintaining the basicity of the apex of the GlpT lumen [52, 170]. However, the "tyrosine-cage" plays an active and direct role in the coordination and binding of the substrate.

Although R269 has been suggested as a part of the putative binding site along with R45 [48, 52, 55, 57, 169] and used as such in a docking study [170], no substrate–R269 interaction was observed in any of our simulations (Fig. 3.2D–F). Despite its structural symmetry to R45, which prompted the assumption that it might play a similar role [48], the simulations invariably show that R269 is not a part of the initial binding site. The distinct roles of R45 and R269 are, indeed, not only supported by mutagenesis experi-

ments [170], but they provide the only consistent interpretation of their results: among the binding site residues (R45, K80, H165, R269), only the mutation of R45 (R45K) resulted in a complete loss of binding and transport in GlpT, whereas R269K still showed binding, although with an increased K_d [170]. R269 might play an indirect role in substrate of binding by contributing to the positive electrostatic potential of the lumen, as manifested by the impairment of P_i^- binding upon neutralization of R269 in System 17. However, it does not seem to be a part of the binding site for the substrate. We also note that our simulations cannot exclude the possibility of direct substrate–R269 interaction during the later stages of the transport cycle. In fact, since R269K mutation [170] is not expected to change the overall electrostatic potential, while still impairing the transport, it is very likely that substrate–R269 direct interaction does become important in later stages of the transport cycle.

3.3.3 Electrostatic Features of GlpT Lumen

The overall positive electrostatic potential generated by the basic residues inside the lumen of GlpT appears to be the key driving force for the observed spontaneous substrate binding, whereas stabilization of the substrate in the binding site requires direct involvement of specific residues as described in the previous section (Fig. 3.2D–F, Fig. 3.7).

The electrostatic potential map calculated using the first 2.5 ns of the simulation of the *apo* system (System 1) is shown in Fig. 3.8. Entrance of a Cl^- ion into the lumen, observed at $t \simeq 2.5$ ns of the *apo* simulation, is consistent with the presence of a strong positive luminal electrostatic potential. The map features a peak at the location of R45 ($\sim$0.76 V), whereas the potential at the location of R269 is not noticeably different from the rest of the protein (Fig. 3.8). The electrostatic potential around R269 is possibly attenuated by E299, which is positioned very close to and forms a salt bridge with R269, while the potential near R45 is amplified by contributions from K46 and K80. Asymmetric distribution of charged residues along the lining of the lumen resulting in the observed electrostatic potential peaking around R45 rationalizes the preference for R45 over R269 in initial substrate binding.

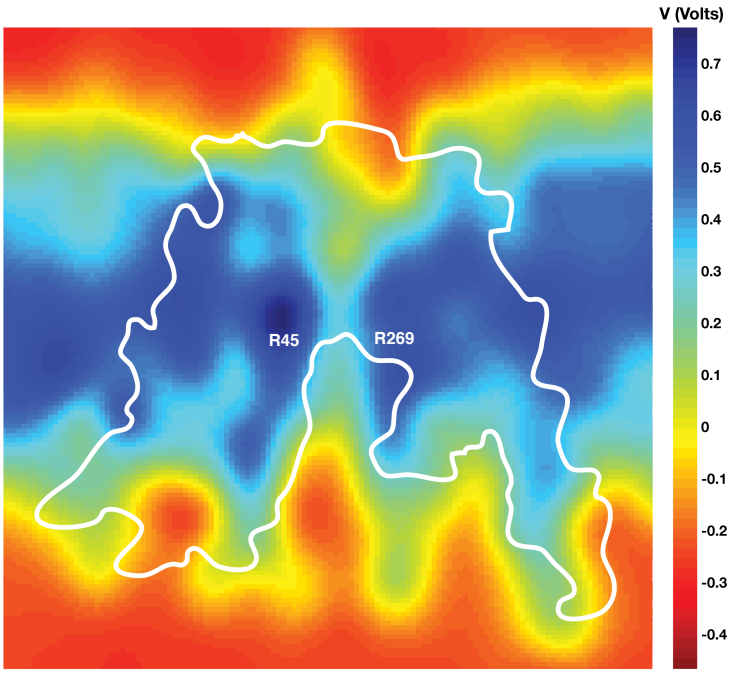

Figure 3.8: Electrostatic potential of GlpT. Electrostatic potential map of GlpT is calculated using the first 2.5 ns of the simulation of the *apo* system. A cross section of the 3D potential passing through the middle of the transporter is shown. The contour of the transporter is shown using the white line, with approximate positions of R45 and R269 labeled. The peak of the potential (dark blue) is located around the position of R45.

The observation that the substrate (P_i^-) did not bind spontaneously after neutralization of either K80, R45, or R269 in our simulations supports the importance of each basic residue for substrate recruitment. We note that, in the case of arginines, charge neutralization was achieved by deprotonation at N_ϵ, in order to preserve the hydrogen bonding capacity of the terminal NH_2 groups. These results indicate that apart from binding the substrate directly, these basic residues play a key role in producing the overall luminal electrostatic positive potential that accounts for the rapid displacement of the substrate from the mouth of the lumen to the binding site.

3.3.4 Partial Closure at Cytoplasmic Half

In all the substrate-binding simulations, a partial but clear closure of the cytoplasmic mouth of the lumen was observed. The extent and the time scale of the partial closure are very similar in all substrate-binding simulations and originate primarily from the

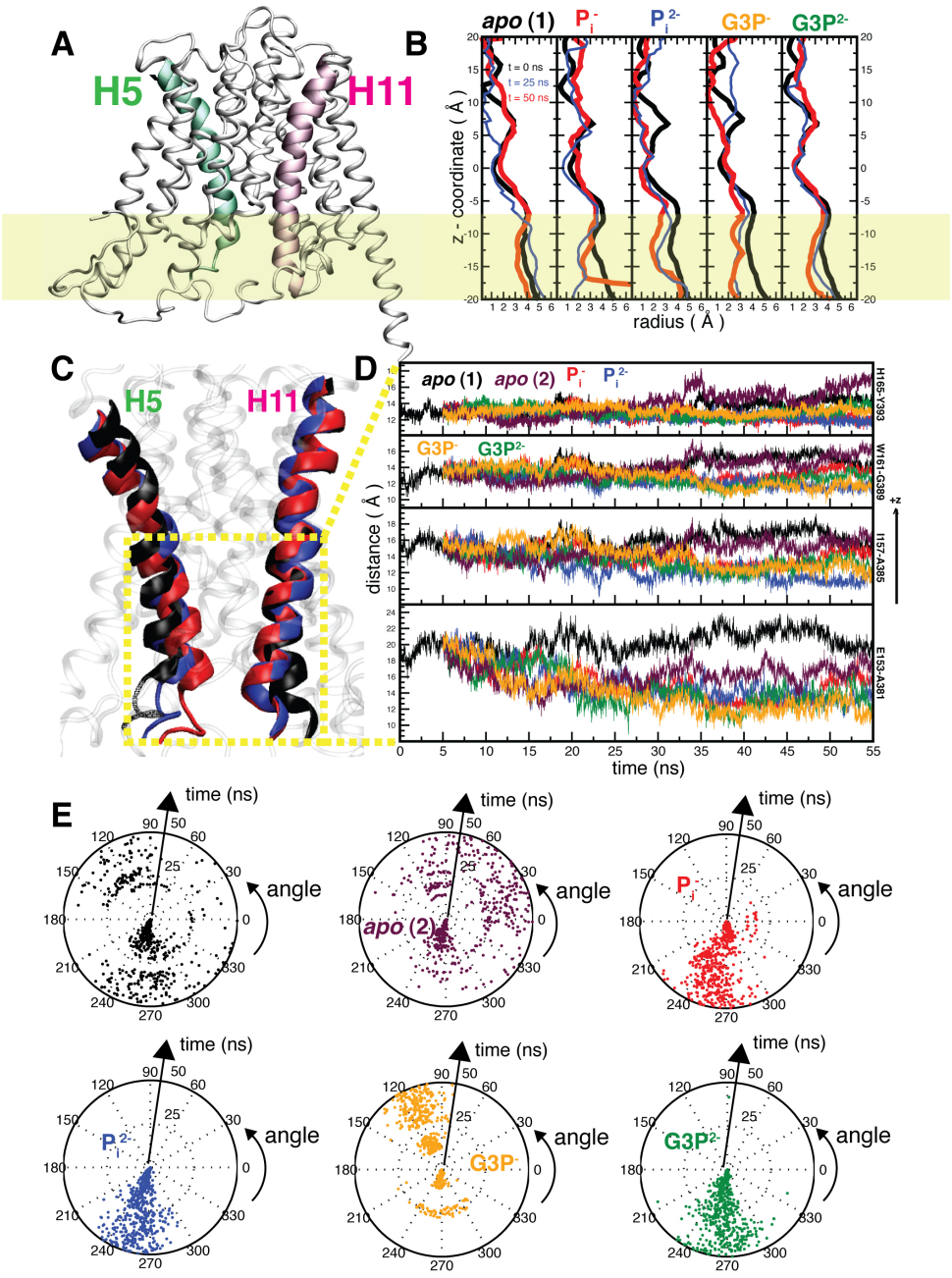

Figure 3.9: Substrate-induced structural changes in GlpT. (A) The crystal structure of GlpT with the two bent helices H5 (green) and H11 (pink) highlighted. The region where the substrate-induced closure takes place is highlighted using a yellow bar. (B) The radius profile of the lumen calculated using HOLE [185]. Substrate-binding simulations show a decrease in radius around the highlighted region. (C) Substrate-induced straightening of H5 and H11 toward the lumen. Black, blue, and red helices represent the structures at t =0 ns, t =25 ns, and t =50 ns taken from the P_i^{2-} binding simulation (System 5). (D) Distances between the C_α atoms of residue pairs on H5 and H11 as a measure of the distance between the two helices. The residue pairs are approximately in the same xy plane. (E) Substrate induced confinement of the rotational degree of freedom of H165. C_α–C_β–C_γ–C_δ dihedral angle of H165 is plotted on a polar coordinate system. The divergence of the data points from the origin represents the dihedral angles at different time steps.

straightening of transmembrane helices 5 and 11 toward the lumen at their cytoplasmic ends (Fig. 3.9). These helices adopt a bent structure in the crystal structure of GlpT [48], and have been reported to exhibit larger flexibility than other helices when simulated individually in membrane [186]. Our simulations now reveal how conformational changes of these two intrinsically highly flexible helices allow them to close the cytoplasmic mouth of the GlpT lumen. In all substrate-bound systems, the closure reaches its maximum around $t \simeq 35$ ns, after which the distance between helices 5 and 11 as well as the size of the lumen stabilize (Fig. 3.9B–D). This phenomenon is consistent with the proposed rocker-switch mechanism and likely represents the initial events along the transition toward the formation of an "occluded" state, in which both the periplasmic and the cytoplasmic sides of the lumen are closed (Fig. 3.1C).

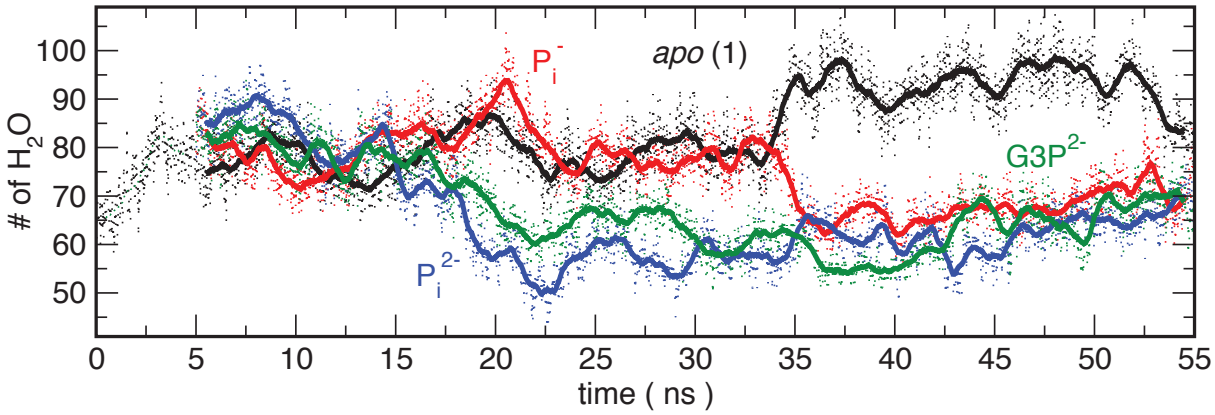

Figure 3.10: Change in the number of water molecules inside the lumen of GlpT. The number of water oxygen atoms within a cylindrical region of cross-sectional radius of 10 Å that covers the lumen of GlpT is calculated. Substrate binding appears to result in a decrease in the number of water molecules.

In contrast to the substrate-bound simulations, in which the partial closure was consistently observed and a partially closed form was maintained throughout the simulations, the *apo* simulations either did not produce this effect (System 1) or resulted in a partially closed state that reverted to the open form toward the end of the simulation (System 2). It appears that the observed partial closure is along a natural soft vibrational mode of the protein, which might occasionally happen even in the absence of the substrate, but is strongly excited by the presence of the substrate. One might also suggest that the substrate stabilizes the conformation (state) resulted from this mode of vibration, as indicated

by the stable partially closed system obtained in all substrate-bound simulations. The observed partial closure can be further quantified by the expulsion of water (Fig. 3.10) from the lumen of GlpT. We note that the partially closed state does not seem to be stabilized by non specific negative ions, such as Cl^- ions, which enter the lumen of GlpT during the simulation of the *apo* system.

The extent of the helical rearrangement is maximal toward the end of the helices 5 and 11, i.e., below the identified binding site (Fig. 3.9), indicating collective motions of these helices, which appear to be stabilized by bridging of several residues by the substrate via hydrogen bonding. Since the phosphate moiety is the common component of all the substrates, the partially closed state is most likely stabilized by the contribution of this moiety. With this regard, the dynamics of H165, a residue located on helix 5, is worth mentioning. In all substrate-bound systems, a clear reduction of the rotational freedom of the side chain of H165 upon substrate binding is observed (Fig. 3.9E). Consistently, substrate binding constrains the rotation around the C_α–C_β–C_γ–C_δ dihedral angle of H165, which is free to rotate in the *apo* systems (Fig. 3.9E). The formation of stable hydrogen bonds between R269 and H165 upon restriction of the rotational freedom of the latter suggests that H165 might act as a "pivot" for bending of helix 5.

3.3.5 Rearrangement of Periplasmic Salt Bridges

Salt bridges on the sealed periplasmic side of GlpT have been implicated as possible switches for the rocker-switch mechanism [48, 52, 170]. This salt-bridge network involves K46 in the N-terminal, as well as D274, E299, and R269 in the C-terminal half. Given that each of K46L, D274N, and E299Q mutants exhibits significantly reduced turnover rates without much loss of binding affinity, the salt-bridge network was suggested to stabilize the cytoplasmic-open state without contributing directly to substrate binding [170]. Our simulations show that substrate binding manipulates the salt-bridge network primarily through affecting K46, which alternates between two essentially immobile residues, D274 and E299 (Fig. 3.11), and, thus, indirectly stabilizes the R269–E299 (Fig. 3.12) salt bridge, an effect that might be related to the confinement of H165 (Fig. 3.9E).

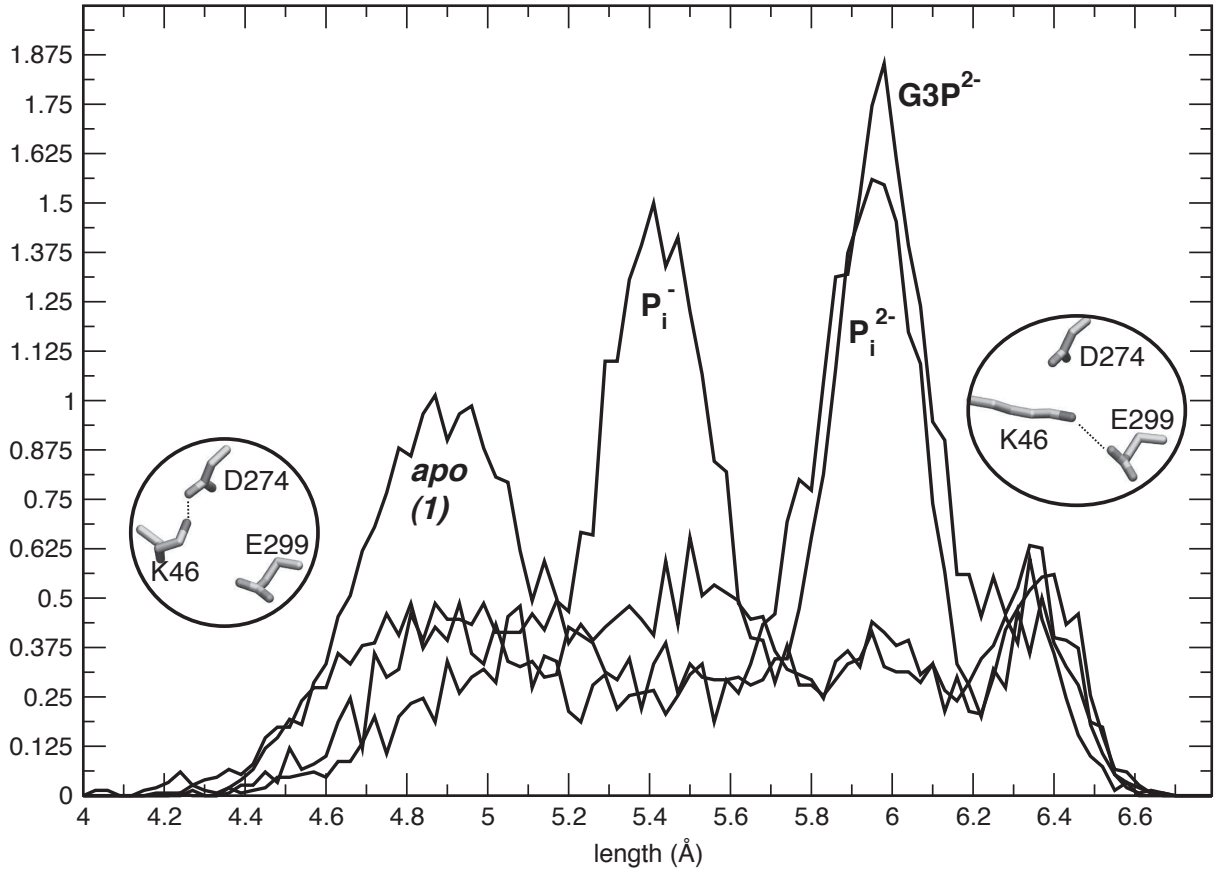

Figure 3.11: Substrate-induced structural changes in GlpT. (A) The crystal structure of GlpT with the two bent helices H5 (green) and H11 (pink) highlighted. The region where the substrate-induced closure takes place is highlighted using a yellow bar. (B) The radius profile of the lumen calculated using HOLE [185]. Substrate-binding simulations show a decrease in radius around the highlighted region. (C) Substrate-induced straightening of H5 and H11 toward the lumen. Black, blue, and red helices represent the structures at t =0 ns, t =25 ns, and t =50 ns taken from the P_i^{2-} binding simulation (System 5). (D) Distances between the C_α atoms of residue pairs on H5 and H11 as a measure of the distance between the two helices. The residue pairs are approximately in the same xy plane. (E) Substrate induced confinement of the rotational degree of freedom of H165. C_α–C_β–C_γ–C_δ dihedral angle of H165 is plotted on a polar coordinate system. The divergence of the data points from the origin represents the dihedral angles at different time steps.

Substrate binding influences the salt-bridge network mainly by forcing K46 into a more extended conformation (longer C_α–N_ζ distance) (Fig. 3.11). While an extended side chain is associated with stabilization of the K46–E299 salt bridge, a compact one favors the K46–D274 salt bridge.

The R269–E299 salt bridge in the N-terminal half, on the other hand, is independent of the K46–E299 salt bridge (Fig. 3.12), but shows a correlation to the hydrogen bonding between H165 and R269. As described above, substrate binding results in structural con-

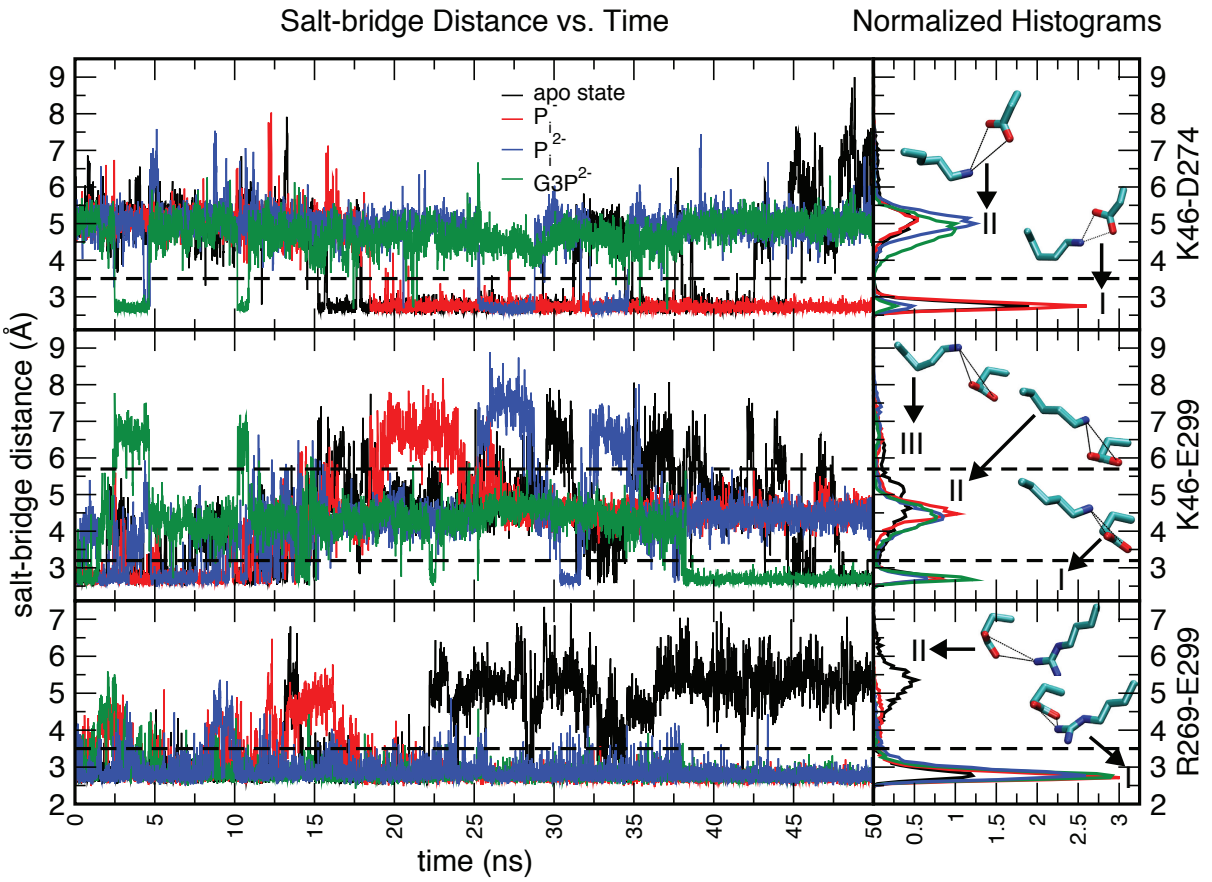

Figure 3.12: Dynamics of periplasmic-side salt bridges. Time evolution of the N–O distance in K46–D274, K46–D299, and R269–D299 (left), along with normalized histograms representing different populations (I, II, III, separated by dotted lines) of salt-bridge lengths (right). The distances represent the minimum distance for all N–O combinations. Representative structures for each population is shown on the right-hand side.

finement of H165 (Fig. 3.9E) in a conformation where the plane of its aromatic side chain is almost parallel to that of the membrane. This conformation promotes hydrogen bond formation between H165 (N_ϵ) and R269, thus, stabilizing the R269–E299 salt bridge. On the contrary, rotational freedom of H165 in the *apo* system (Fig. 3.9E) leads to destabilization and subsequent rupture of the R269–E299 salt bridge (Fig. 3.12). Salt-bridge reorganization appears to constitute an important set of molecular events facilitating the rocker-switch type conformational change of the transporter, resulting in weakening of the interactions on the periplasmic interface of N- and C- terminal halves [187]. Similar phenomena have also been observed in other simulation studies upon substrate binding [188].

3.4 Summary

Extended equilibrium MD simulations were performed on the glycerol-3-phosphate transporter, GlpT, in the presence of its natural substrates (P_i or G3P) at various physiologically relevant protonation states. The simulations successfully capture spontaneous binding of the substrates, characterize the pathway of substrate translocation inside the lumen and the binding site, and reveal initial protein conformational changes induced by substrate binding.

The binding process involves K80 and R45, with the former acting as a "hook" that recruits and steers the substrate into the binding site, and the latter acting as a "fork" tightly holding onto the phosphate moiety of the substrate in the binding site. In the bound state, the phosphate moiety is stabilized further by H165 and by several surrounding tyrosine residues (Y38, Y42, and Y76) forming a cage-like binding pocket. No direct interaction between the substrate and R269 is observed within the time scale of our simulations, indicating its distinct function in initial substrate binding despite its structural symmetry relation to R45.

Calculation of the luminal electrostatic potential revealed that the potential peaks approximately at the position of R45, explaining its preferential substrate binding over R269. Contributions to the luminal electrostatic potential by R45, K80, and R269 appear to be essential for initial substrate recruitment, since neutralization of either residue impairs binding.

Our simulations also reveal significant substrate-induced global conformational changes of GlpT consistent with the rocker-switch model. Substrate binding was found to trigger partial closing of the cytoplasmic opening of GlpT, likely constituting the initial steps toward the formation of an "occluded" state (IF-occ). Besides, substrate binding alters the salt-bridge interactions in the periplasmic half, an effect that prepares the protein for opening on this side in order to form the periplasmic-open state during the transport cycle in GlpT.

Chapter 4

Structural Basis of Substrate Selectivity in GlpT *

4.1 Introduction

Major Facilitator Superfamily, as a whole, transports an enormous diversity of small polar or charged species, including ions, sugars, organic phosphates, drugs, neurotransmitters, amino acids, and even peptides, although its evolutionarily related individual members transport only one or a few related substrates. How can individual family members generally display stringent specificity toward the substrates they transport while at the same time the entire family recognizes and transports such a myriad of substrate types? The substrate specificity of each transporter must therefore be defined by only a few amino acid residues and, in particular, by differences between primary sequences at the substrate binding site [55].

Given that the structures of all MFS transporters solved to date are similar [48, 61–63, 189–193], indicating that all the family members are probably architecturally alike, it remains to be determined how each MFS transporter differentiates its cognate substrate from an array of substrates that are often chemically and structurally very similar. Pre-

*This chapter appeared as a research article in Biophysical Journal [5]. Christopher J. Law, Giray Enkavi, Da-Neng Wang, and Emad Tajkhorshid, "Structural basis of substrate selectivity in the glycerol-3-phosphate:phosphate antiporter GlpT". *Biophysical Journal*, 97:1346–1353, **2009**.

vious descriptions of the basis of substrate selectivity in several transporters have been hampered by a lack of high-resolution structures [194–201]. In contrast, studies on the MFS lactose/H+ symporter LacY, for which high-resolution structural information is available [191, 193], have enabled the amino acid residues involved in substrate specificity and binding to be well documented [202–207]. Recently, new MFS transporters with bound substrates have been crystallized, revealing the amino acid residues involved in binding [62, 63]. Despite this, a detailed description of the structural basis of substrate selectivity is lacking for many other MFS transporters.

Residues in GlpT that bind to P_i or the phosphate moiety of G3P when the transporter is in the IF conformation were previously identified to be R45, K80, H165, and R269 (Fig. 3.1B) [48, 170]. The affinity of the transporter for G3P is much higher than it is for P_i [60, 167], and this discrimination is crucial for the transporter function in that the rate of transport is determined by substrate association/dissociation reactions [52, 169]. When the substrate-binding site is exposed to the periplasm (the OF conformation), the much lower affinity of the transporter to P_i allows that substrate to be replaced by G3P. After being transported across the membrane into the cytoplasm, G3P is released and replaced by P_i owing to the much higher intracellular concentration ($\sim$4 mM) of the latter [208], thus perpetuating the transport cycle.

We hypothesized that only a few of the residues that line the lumen of GlpT are required to impart discrimination between substrates. To test this hypothesis, we performed MD simulations to simulate the process of binding of P_i and G3P to GlpT in the IF conformation, and to probe residues that may be involved in substrate recognition and selectivity. Subsequent biochemical analyses using binding assays of GlpT mutants in detergent solution and transport assays on reconstituted proteoliposomes were used to test the function of each selected residue by our collaborators, lab of Da-Neng Wang in New York University. This approach has enabled us to offer a description of the structural basis of substrate selectivity in this integral membrane antiporter.

4.2 Methods

4.2.1 System Preparation

The GlpT structure [48] (PDB ID 1pw4) was used as the initial model, in which the mutated residues were reverted back to the wild-type ones, and missing side chains were modeled using the PSFGEN plugin of VMD [164]. The missing interdomain loop was also modeled, initially as an unstructured chain, and relaxed in vacuum for 100 ps with the rest of the protein fixed. After internal water molecules were added with DOWSER [175], the protein was embedded in a patch of POPE (1-palmytoil-2-oleoyl-sn-glycero-3-phosphatidylethanolamine) bilayer with the membrane normal along the z axis. The lipid molecules overlapping with the protein were removed, and the system was solvated and ionized with 100 mM NaCl to produce an electroneutral system. The final dimensions of the system were ($115 \times 115 \times 11$ Å^3), including ∼125,000 atoms. Then, the system was subjected to a series of energy minimization and partially constrained MD relaxations before it was freely equilibrated for 5 ns under NP_nT) (constant area) conditions. The resulting system (Fig. 3.1A) was used as a starting point for all subsequent simulations, unless specified otherwise.

4.2.2 Simulation Systems

The substrates (P_i and G3P) were initially placed close to the mouth of the lumen on the cytoplasmic side of GlpT, with the phosphate moieties ∼15 Å away from R45 (the putative binding site). After minimization for 5000 steps was completed, each system was simulated for 50 ns to establish stable binding of the substrates and obtain a clear description of the binding site as the substrates explored it. H165 was maintained in an unprotonated state during the simulations. Additionally, GlpT was also simulated in its *apo* form as a control. The W138R mutant was simulated both in the *apo* state and in the presence of either P_i or G3P.

4.2.3 Simulation Protocols

Simulations were carried out using a time step of 1 fs, a temperature of 310 K, and a pressure of 1 atm along the z-direction (NP_nT). The simulations were performed with NAMD 2.6 [163] using the CHARMM27 force field with ϕ/ψ cross-term map (CMAP) corrections [161]. Water molecules were modeled as TIP3P [179]. The force-field parameters for the substrates were adopted from similar molecules in the CHARMM force field. Constant pressure was maintained by the Nos-Hoover Langevin piston method [180, 181], and constant temperature was maintained by Langevin dynamics with a damping coefficient of 1 ps^{-1} for heavy atoms. The short-range interaction cutoff was set to 12 Å. Long-range electrostatic interactions were computed using the particle mesh Ewald (PME) method (46) with a grid density of at least 1 $Å^{-3}$. Bonded, non-bonded, and PME calculations were performed using 1, 2, and 4 fs time steps, respectively.

4.2.4 Specific Analysis

To characterize the interaction between the protein and the substrate, a contact analysis was performed on the 50-ns-long substrate-binding trajectories. All residues with any atom within 3 Å of either the phosphate or the glycerol moiety (in the case of G3P) were considered interacting. The probability of contact between the substrate and the protein was then defined as the ratio of the number of trajectory frames in which a contact existed over the total number of trajectory frames. To filter out transient contacts, only residues exhibiting contact probabilities higher than 25% for the phosphate moiety and 15% for the glycerol moiety are reported.

4.2.5 Summary of Experimental Methods

Our experimental collaborators, Da-Neng Wang Lab (New York University, NY) expressed wild-type and mutant GlpT proteins, purified, and reconstructed them into the proteoliposomes. They, also, performed transport assays and substrate binding affinity assays using quenching of the intrinsic tryptophan fluorescence. For details of the experimental

methods, see reference [165]

4.3 Results and Discussion

4.3.1 Identification of the substrate specificity determinants

To investigate the molecular basis of differences in binding mode and affinity of GlpT for organic and inorganic phosphates, we first studied transporter-substrate interactions by means of MD simulations. Using the crystal structure of GlpT in the IF conformation [48] as the starting point (Fig. 3.1A), we were able to describe spontaneous binding of the substrates to the transporter without applying any biasing potential or force. These unbiased simulations were performed in the presence of either P_i or G3P initially placed $\sim$15 Å away from R45 in the putative binding site. Both substrates were simulated in their divalent form to exclude complications due to the effect of titration state on the results. Although it is not known whether or to what degree the pK_a of the substrates is affected upon entrance into the lumen, the divalent forms of the substrates appear to be more relevant to the transport cycle of GlpT.

Therefore, we will focus our discussion mainly on the results obtained for divalent species, noting that the use of monovalent species did not affect our conclusions. The simulations revealed a rapid, spontaneous translocation of the substrates (in both cases) from their initial position at the mouth of the lumen to its apex, where they stayed bound for the remainder of the simulation. The trajectories allowed us to differentiate the protein residues that specifically interact with either substrate from those that contribute to the binding of only one. We note that we, also, tested the effect of all different physiologically relevant titration states of P_i and G3P and the results are presented in Chapter 3.

The binding trajectories allowed us to quantify and compare the contact patterns of dibasic P_i and G3P with the protein (Fig. 4.1). Of importance, a direct comparison of the two substrates reveals distinctive steric and electrostatic effects arising from the glycerol moiety of G3P that may play a role in its higher binding affinity. Because GlpT transports both G3P and P_i but not glycerol [167], we hypothesized that interactions

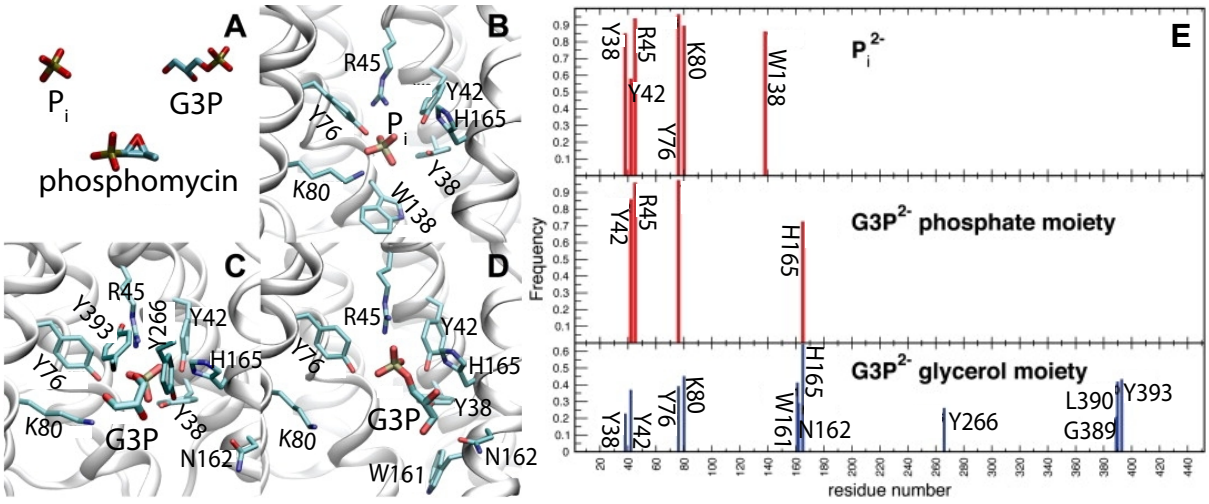

Figure 4.1: Substrate binding in GlpT. (A) Molecular structures of GlpT substrates, dibasic phosphate (P_i^{2-}), dibasic glycerol-3-phosphate ($G3P^{2-}$), and phosphomycin. (B–D) The protein is rotated ∼90° around the z axis with respect to the view in Fig. 3.1A to enable visualization of all the binding-site residues. (B) Binding mode of P_i^{2-}. The phosphate group establishes close electrostatic interactions with R45 and K80 while forming hydrogen bonds with Y38, Y42, Y76, and H165. (C) First binding mode for G3P. The phosphate moiety of G3P establishes interactions with the GlpT binding site similar to those in the case of P_i. The backbone of G3P forms a ring-like structure (stabilized by intramolecular hydrogen bonds between the hydroxy and phosphate groups) oriented parallel to the plane of the membrane and forms a stacking interaction with Y393, and hydrogen bonds with K80, Y76, and Y266. (D) Second binding mode of G3P. The phosphate moiety is in the same location as before and forms the same interactions with the binding pocket. The backbone of G3P is in an extended conformation approximately parallel to the membrane normal, forming hydrogen bonds with N162. (E) The results of the contact analysis between the substrate and GlpT (see 4.2). A cutoff distance of 3 Å between any atom from the protein and the substrate is used to define residues that make contact with the substrate. Only residues that maintain contact with the substrate during a fraction >25% (for the phosphate groups) and 15% (for the glycerol moiety of G3P) of the duration of the simulation are shown.

between GlpT and the glycerol moiety of G3P hold the key for understanding the protein's substrate specificity. The transporter-substrate interactions can therefore be broken into two main components: the interaction of GlpT with the phosphate moiety based on the data obtained from both the P_i^{2-} and $G3P^{2-}$ simulations, and those involving the glycerol moiety as identified in the $G3P^{2-}$ simulation. During 50-ns simulations, the residues involved in recognition of P_i include the positively charged side chains of R45 and K80 as well as Y38, Y42, Y76, and H165 (Fig. 4.1B); interactions with Y76, R45, K80, and Y38 occur with the highest frequencies (Fig. 4.1E). As expected, the interactions involving these residues are mostly electrostatic. The same set of residues contribute to the recognition of both P_i and the phosphate moiety of G3P, although the frequency of

substrate contacts for some of these residues (e.g., Y38, K80, and W138) varies from one substrate to another, a difference that is likely due to limited sampling rather than fundamental differences in the binding mode of the phosphate groups in these species. A similar binding mode can be also expected for the phosphate moiety of other small organophosphates, such as phosphomycin. The results indicate that K80, R45, and H165 form the core of the binding site for the phosphate moiety. This is in accord with previous results showing that mutation of K80 or R45 to alanine and lysine, respectively, killed heterologous G3P-P_i transport of the protein reconstituted into proteoliposomes, and mutation of H165 resulted in a transport rate that was only ∼6% of that catalyzed by wild-type protein [170].

4.3.2 Glycerol-3-Phosphate (G3P) Binding Modes

Several GlpT residues were found to interact with the glycerol, but not the P_i moiety of G3P. Two binding modes, which keep the GlpT-phosphate interaction invariant but differ in their GlpT-glycerol interaction, were observed and characterized during the 50 ns simulation of G3P binding (Fig. 4.1C and D). The first binding mode is apparent during the first half of the simulation and lasts for ∼20 ns. In this mode, the glycerol backbone lies parallel to the plane of the membrane, making hydrophobic contacts with the side chain of Y393, whereas the terminal carbon atom of the glycerol backbone is in contact with the backbone of G389 and L390 (Fig. 4.1C). The binding is further stabilized by hydrogen bonds between the hydroxy groups of glycerol and the K80 and Y266 side chains. Here, the backbone of G3P adopts a ring-like structure, in which internal hydrogen bonds between the hydroxy groups and the phosphate moiety of G3P form an arrangement that might be favorable due to lack of sufficient hydration for the hydroxy groups in this region. This conformation, which is somewhat reminiscent of the cyclic structure of phosphomycin (Fig. 4.1A), is ideal for optimal interaction with Y393. In the second binding mode, which is captured during the second half of the simulation and lasts for ∼25 ns, G3P adopts a more extended conformation and is no longer parallel to the plane of the membrane (Fig. 4.1D). It loses interaction with the K80 side chain, but forms

contacts with the side chain of H165 as well as hydrogen bonds with N162. There is also a contact between the substrate and the backbone and the C_β atom of W161 in this configuration. Other tyrosine residues (Y38, Y42, and Y76) that line the GlpT lumen also show hydrogen bond interactions with G3P and P_i. Because of its larger size, however, G3P is the only substrate that can also interact, via its glycerol moiety, with the side chains of N162, Y266, and Y393 while bound to R45. Besides the interactions with the GlpT backbone and the residues that also interact with the phosphate moiety of G3P, our simulations suggest a role for the N162, Y266, and Y393 side chains in selecting between P_i and G3P when GlpT is in the IF conformation.

4.3.3 Experimental Validation of G3P Binding Site

Table 4.1: Heterologous G3P-P_i exchange transport activity and apparent binding dissociation (K_d) constants of P_i, G3P, and phosphomycin binding to wild-type and mutant GlpT

GlpT protein	**K_d of P_i (μM)**	**K_d of G3P (μM)**	**K_d of phosphomycin (μM)**	**G3P–P_i exchange transport activity**
Wild-type	7.4 ± 0.4	0.8 ± 0.2	0.18 ± 0.02	Yes
N162A	4.4 ± 1.4	No binding	No binding	Not Determined
Y266F	1.7 ± 0.4	No binding	No binding	No
Y393F	8.6 ± 0.7	No binding	43 ± 6	Not Determined

To confirm that the side chains of N162, Y266, and Y393 are indeed directly involved in binding to the glycerol, but not the phosphate moiety of G3P, these three residues were mutated by our experimental collaborators to alanine, phenylalanine, and phenylalanine, respectively. The effects of these GlpT mutations both on the binding affinity of the protein to each substrate and the G3P-P_i exchange transport activity of each mutant were, then, measured. The N162A, Y266F, and Y393F mutants bound P_i in detergent solution at neutral pH with apparent K_d values that are very similar to the K_d measured for wild-type GlpT binding to P_i (Table 4.1). thus demonstrating that these residues do not interact with P_i. In contrast, none of the above mutants showed any binding to G3P under the assay conditions (Table 4.1), suggesting that each of these residues plays an important role in the binding and recognition of G3P, particularly its glycerol moiety. It should be

noted, however, that G3P may still bind very weakly to the GlpT N162A, Y266F, and Y393F mutants, since the binding assay can only measure K_d values up to hundreds of micromolars concentration range. Moreover, under the assay conditions, a molecule of DDM is bound to the N-terminal half of the lumen of the transporter, stabilizing the IF conformation [170] and must be competed off before the substrate can bind. The measurements also reflect a composite of binding to both IF and OF conformations, but with the major contribution arising from substrate binding to transporter in the IF conformation. Thus, whether there is a difference in the binding constants for the same substrate in the different transporter conformations, or indeed whether different residues are involved remains to be investigated.

No heterologous G3P-P_i exchange transport activity was observed for the Y266F mutant (Table 4.1) as expected, since both substrates must be able to bind tightly to the protein for transport to be catalyzed. As a result, transport assays were not performed on the other two mutants. Thus, the binding and transport assays of the site-directed mutants, along with MD simulations, clearly showed that N162, Y266, and Y393 indeed play a role in discriminating between organic and inorganic phosphates.

The fact that mutation of residues associated with either of the two G3P binding modes identified in the simulations results in deleterious effects on substrate binding underlines the relevance of each binding mode. Although the two sets of G3P-interacting residues could represent sites that are sampled by the substrate at various time points during the initial stages of binding, we favor an alternative explanation. Examination of the crystal structure of GlpT [48] shows that these two sets of G3P-interacting residues identified in the simulations are somewhat spatially apart, and thus a small substrate such as G3P is unlikely to simultaneously interact with both sites in the absence of large protein conformational changes. This agrees with the suggestion that G3P binding results in the GlpT N- and C-terminal domains moving closer together to form an occluded state [60, 169], with a subsequent spatial convergence of these residues and the formation of a common binding site that allows simultaneous interaction of G3P with the side chains of N162, Y266, and Y393. The apparent loss of binding affinity to G3P upon mutation of

these residues at the substrate concentrations suggests that the contribution to binding affinity from these residues is not just additive, but the presence of each residue is a requirement for binding of G3P. According to our binding model, any of the residues that interact with a particular substrate are necessary for tight binding of that substrate. Moreover, we suggest that mutation of any of these residues renders the binding site unfavorable for that substrate, possibly by altering the local geometry and properties of the binding site.

4.3.4 Phosphomycin Binding Site

Elucidation of the GlpT substrate-specificity determinants derived from the GlpT-G3P and GlpT-P_i interactions immediately suggests that the antibiotic phosphomycin, whose structure can be regarded as intermediate between that of P_i and G3P (Fig. 4.1A), will interact with some of the same residues (more residues than does P_i, but fewer than G3P). This was also tested by binding experiments. While the N162A and Y266F mutants showed no binding to phosphomycin, the Y393F mutant did bind it (Table 4.1). This is ∼240-fold less tight than binding of the wild-type transporter to phosphomycin (Table 4.1)). These studies provide strong support for our prediction that N162 and Y266 play an important role, probably in recognizing the phosphomycin epoxide group as well as the glycerol moiety of G3P.

4.4 Summary

The structural basis of substrate selectivity in GlpT in the IF conformation depends on only a handful of residues that line the transporter's lumen. These residues can be divided into three groups: 1), the side chains of R45, Y38, Y42, and Y76 that recognize and bind only P_i and the phosphate moiety of G3P; 2), the K80 and H165 side chains that can interact with both P_i and the glycerol moiety of G3P; and 3), the N162, Y266, and Y393 side chains and the backbones of G389 and L390 that recognize and bind only the glycerol moiety of G3P. Interactions with the second and third groups of residues give

G3P a higher affinity than P_i to GlpT. In the outward-facing OF conformation, some of the residues responsible for the protein's substrate specificity may be different. We believe that these results may imply a common mechanism for the substrate diversity among the entire MFS, and specificity of each member.

Chapter 5

Formation of Transient Channel-like States in GlpT*

5.1 Introduction

Various experimental studies have reported uncoupled cotransport of ions and water for a number of membrane transporters [210–212]. Various models have been proposed to explain this phenomenon [210]. A number of models propose that the substrate gains access to its binding site along with several water molecules which are then occluded from other side of the membrane together with the substrate [213]. Other models suggest that the water cotransport is driven by local hyperosmolar conditions created due to the accumulated substrate [214–216]. The former mechanism, also referred to as the carrier-mediated model [216], is in agreement with a perfectly coordinated alternating access cycle, while the latter suggest that water or other species may leak through the transporter. While the two views differ substantially, either one has received support from experimental data on water cotransport [214, 215, 217, 218].

Although no study implicating GlpT in water transport currently exists, other MFS

*This chapter appeared as a research article in Proceedings of the National Academy of Sciences [209]. Jing Li, Saher A. Shaikh, Giray Enkavi, Po-Chao Wen, Zhijian Huang, and Emad Tajkhorshid, "Transient formation of water-conducting states in membrane transporters". *Proceedings of the National Academy of Sciences, USA*, 110(19):7696–7701, **2013**.

transporters, most prominently glucose transporters GLUT1 and GLUT2, have been reported to exhibit water transport [219,220]. In this study, we performed extended simulations of GlpT and report the formation of channel-like intermediates permeable to water. Based on the characterized water conducting states we reported for transporters from a diverse set of families [209], we argue the channel-like states that result in uncoupled water transport may, indeed, be a universal aspect of the alternating access mechanism.

5.2 Methods

5.2.1 System Preparation

For detailed description of modeling GlpT in the membrane, and simulation protocols refer to Section 3.2.2. The simulations described here were prepared as follows: The initial coordinates were taken from t=50 ns frame of previously reported spontaneous substrate (G3P) binding simulation [5, 165] and H165 was protonated.

5.2.2 Simulation Systems

The protonated system was subjected to 5000 steps of minimization, followed by 200 ns equilibrium simulation.

5.2.3 Simulation Protocols

The simulations were performed with NAMD 2.6 [163] using the CHARMM27 force field with ϕ/ψ cross term map (CMAP) corrections [161]. All the production simulations were carried out using a 2-fs time-step, at a constant temperature of 310 K, and a constant pressure of 1 atm maintained only along the z direction, i.e., normal to the membrane (NP_nT). Water molecules were modeled as TIP3P [179]. The force field parameters for $G3P^{2-}$ was adopted from similar molecules ($Me{-}PO_4^{2-}$, and glycerol) in the CHARMM force field. Constant pressure was maintained by the Nosé-Hoover Langevin piston method [180,181], and constant temperature by Langevin dynamics with a damping coefficient of $0.5\,ps^{-1}$

for non-hydrogen atoms. Short-range interaction cutoff was set to 12 Å. Long-range electrostatic interactions were computed using the particle mesh Ewald (PME) method [182] with a grid density of at least 1 $Å^{-3}$. Bonded, non-bonded, and PME calculations were performed at 2-, 2-, and 4-fs intervals, respectively.

5.2.4 Specific Analysis

In order to quantify water permeation, full permeation events were counted, i.e., a water molecule has to traverse the entire membrane span in order to be counted as a permeation event. This is done by defining two planes parallel to the membrane and positioned at different points along the membrane normal, with one on the cytoplasmic side and the other on the periplasmic/extracellular side in a way that they cover the entire lumen region of the transporter. The planes are about 30 Å apart. For a water permeation to be counted, a water molecule needs to have crossed the membrane and both planes, i.e., traveling from one side of the membrane all the way to the other side. The analysis is confined to a cylindrical region covering the lumen of the protein, and therefore excluding any possible water leak through the lipid bilayer.

5.3 Results and Discussion

5.3.1 Formation of Channel-Like State

During a 200 ns simulation of substrate-bound, IF-state GlpT [5,165], we observed copious water transport events (Fig. 5.1). Water permeation in GlpT occurs via an aqueous pathway running through the substrate binding site and the radius of the constriction point remains smaller than that of the main substrate (Fig. 5.1). Water permeation in GlpT was only observed following global conformational changes in the protein structure. Starting from the crystal structure in *a*po IF-state [48], GlpT attains the observed water-conducting state only after two mechanistically relevant events deemed to be necessary for the IF↔OF transition, namely, substrate binding and protonation of a histidine located

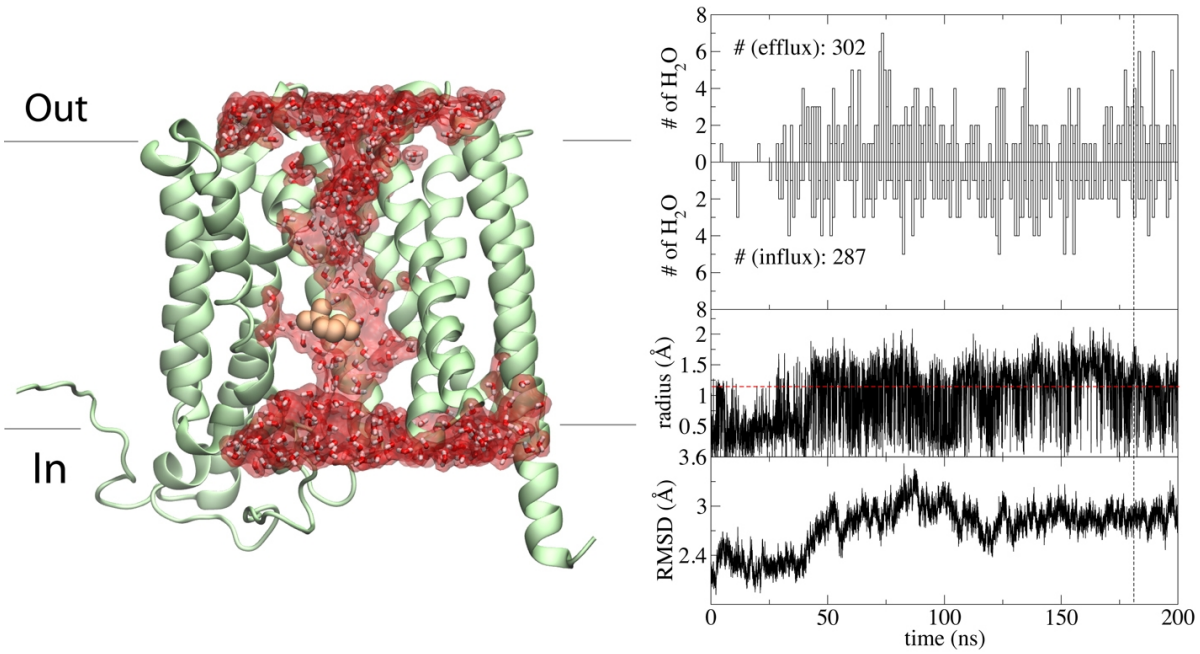

Figure 5.1: Water conduction through GlpT (Left) Representative water-conducting frame from the simulations. (Right) Time series for the number of water permeation events along the efflux and influx directions, radius of the narrowest part along of the aqueous lumen, and C_α-RMSD of the transmembrane region. The vertical black dashed lines represent the snapshots shown in the molecular images, and the horizontal red dashed lines show the minimal radius to accommodate a water molecule.

in the vicinity of the bound substrate [52, 170].

5.3.2 Conformational Changes Leading to Channel-Like State

The first event is spontaneous substrate binding (Fig. 5.2) and subsequent partial occlusion of the cytoplasmic side of GlpT ($\sim$0.8 Å increase in C_α RMSD of the transmembrane helices, Fig. 5.2) (see Chapter 3). The partially occluded state of GlpT obtained in these simulations is hermetically sealed to water and substrate permeation on the periplasmic side. On the other hand, it is closed sufficiently on the cytoplasmic side, preventing back diffusion of the substrate while being accessible to water.

The second event is the protonation of the binding site histidine (H165) located in the vicinity of the bound substrate. The protonation of H165, in response to changes in the local electrostatic environment of the binding site due to the proximity of the charged substrate, is considered as one of the key events triggering IF$\leftrightarrow$OF transition [52, 170]. Previous computational and experimental studies established a coupling between

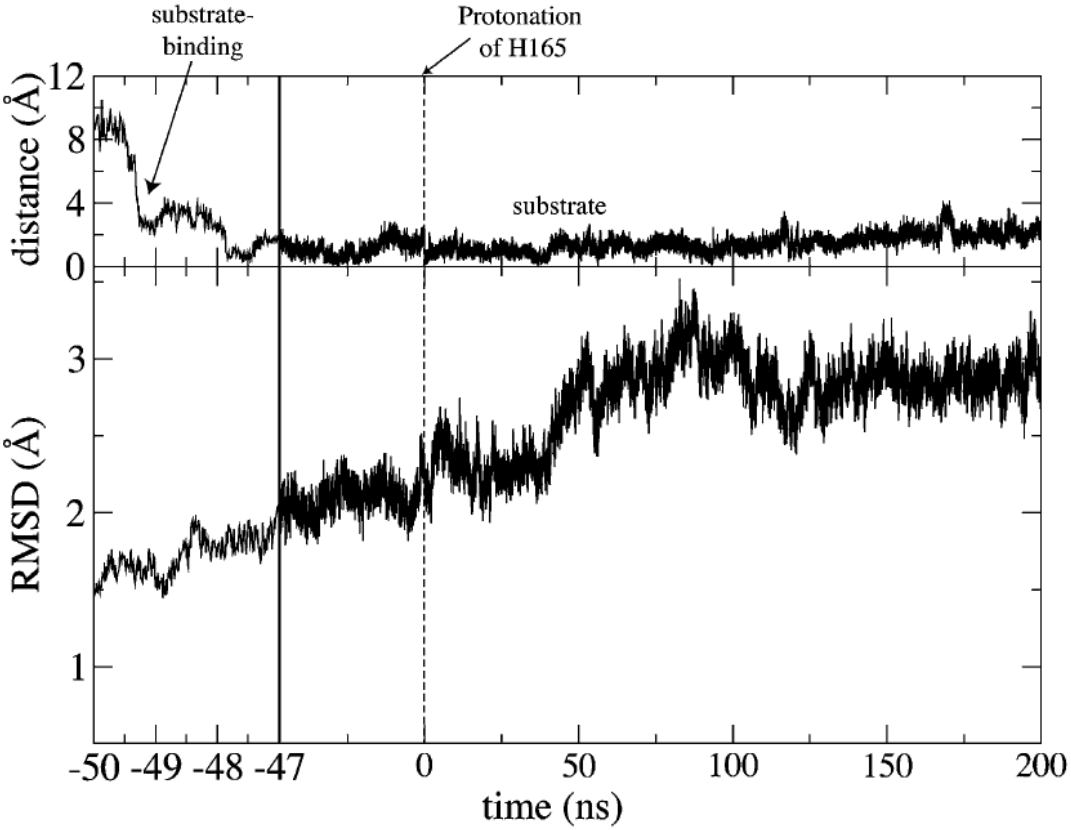

Figure 5.2: Substrate binding and protonation of H165 result in conformational change in GlpT. Top panel: The distance of phosphorus atom of $G3P^{2-}$ from its initial position (t=-50 ns), ∼10 Å away from the binding site of GlpT (defined as the position of the phosphorus atom of $G3P^{2-}$ at t=0 ns) as a function of time. The bottom panel shows the C_α-RMSD of the transmembrane helices as defined in [48] with respect to the crystal structure. First 3 ns (-50 ns – -47 ns) of the simulation is shown in larger scale to show the substrate binding event clearly. Dashed line (t=0 ns) indicates H165 protonation at N_ϵ. Largest structural change happens early on as the protein relaxes in the membrane. The main conformational change in response to substrate binding happens only on two helices between t=10 ns and t=50 ns with little effect on the overall RMSD. Protonation of H165 at t=∼0 ns results in a conformational change which is reflected in the RMSD.

the protonation state of H165 and the dynamics of the salt bridges that are believed to maintain the sealed state of the periplasmic side [52,170]. H165 is likely to accept an extra proton only upon changes in the local electrostatic environment of the binding site due to the proximity of charged substrate, namely, substrate binding may increase the intrinsic pK_a of H165 and result in temporary protonation of H165. We have calculated the pK_a of H165 at 5 ns intervals from our substrate binding simulations reported in Chapter 3. The calculations show that substrate-bound simulations show appreciable increase in the intrinsic pK_a of H165 (Fig. 5.3).

We, then, tested the effect of the protonation of H165 in a 200 ns long simulation starting with the $G3P^{2-}$-bound GlpT; that is, the final state obtained from the $G3P^{2-}$-binding simulation (see Chapters 3 and 4) [5, 165] (Fig. 5.2). The resulting structural changes on the periplasmic side triggered the separation of the helices that seal the periplasmic side ($\sim$1 Å increase in C_α RMSD of the helices, Fig. 5.2). While the transporter remains occluded to the substrate, the separation is large enough to form a water-conducting state, one that is occluded to the substrate but permeable to water from either side of the membrane. These results offer a putative mechanism of water transport for other MFS members as well, in particular for GLUT1 and GLUT2, where water transport has been demonstrated experimentally [219, 220].

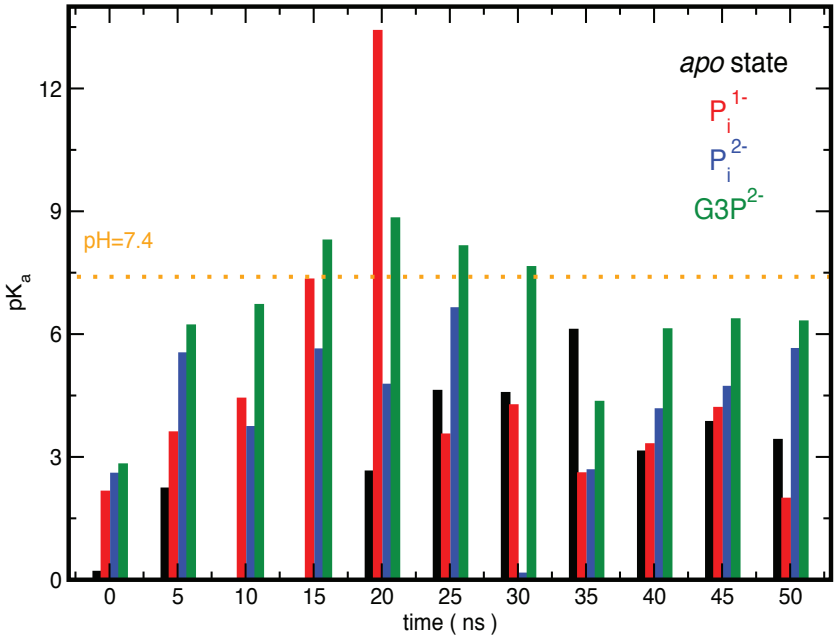

Figure 5.3: pK_a of H165 calculated from snapshots taken at 5 ns intervals from simulations by H++ [221]

5.4 Summary

Earlier experimental studies on several transporter families have detected conduction for water and/or ions [210, 211, 219, 220, 222–225]. Our simulations suggest that GlpT is prone to formation of channel-like intermediate states during its IF↔OF conformational

Table 5.1: Simulated systems

Transporter	Substrate	Family or superfamily	State	Simulation time
vSGLT	Na^++/glucose	SSS	IF substrate-bound	1.15 μs
Glt_{ph}	Na^++/glutamate	SLC1	intermediate *a*po	160 ns
GlpT	glycerol-3-phosphate/P_i	MFS	IF substrate-bound	200 ns
Mhp1	Na^++/benzyl-hydantoin	NCS1	OF substrate-free	1.2 μs
maltose transporter	maltose	ABC	OF	100 ns

transition, resulting in water conduction. We observed similar water-conducting channel like states, also, in a diverse set of other membrane transporters (Table 5.1) [209]. Our simulations provide a mechanistic/structural explanation to the water-conducting states reported in experimental studies and suggest that opening and closing motions of the two gates in the alternating access mechanism might not be perfectly controlled. Water transport is mediated by well characterized water-conducting channel-like states most likely during the transition between major functional states. In most of the cases, the aqueous pathway forms along the putative substrate pathway. Moreover, our results demonstrate that these channel-like water conducting states are not necessarily detrimental to the alternating access mechanism, since the channels remain narrow enough to prevent leaking of the substrate. In other words, while the transient water-conducting states can be readily permeable to water, or even ions, they remain effectively occluded to the substrate, hence not expected to hinder the uphill transport function of the transporter. Although based on our data, we cannot discard the involvement of an active (carrier-mediated) mode of transports [210], our results favor that water transport occurs via channel-like states [214–216].

Chapter 6

Thermodynamic Characterization of Transport Cycle in GlpT *

6.1 Introduction

In our earlier work, we have concentrated on the equilibrium dynamics of the IF state GlpT in a physiologically relevant simulation setting [2, 5, 165, 209, 226–228]. Our simulations helped us characterize the structural determinants of substrate selectivity, and identify relevant binding modes of substrates. We were also able to describe conformational changes that are initiated by substrate binding and can be interpreted as the partial formation of the IF-occluded state of GlpT. Besides, based on our extended simulations, we could even capture a "channel-like" state of GlpT. Despite this water-conducting state at first seemed incongruent with the "alternating access mechanism", under the light of experimental data and a comprehensive simulational studies of many transporters, we now believe that "water-conducting" states are indeed a natural consequence of the alternating access mechanism.

Our results, however, were merely based on simulations of the IF state of the GlpT. Despite, a wealth of information provided by long and extensive simulations, it is practi-

*This chapter is in preparation as a research article. Mahmoud Moradi, Giray Enkavi and Emad Tajkhorshid, "Thermodynamic Characterization of Transport Cycle in GlpT".

cally impossible to capture and characterize the transport cycle of any transporter within the accessible time scales of MD. Modeling the OF state GlpT and characterization of the IF↔OF pathway, have been our long standing interest, which was inhibited by the lack of structural information for any MFS transporter until FucP was characterized. We had been able to capture functional modes based on anisotropic network model (ANM) that could lead to an OF-occluded state [229–232], but failed to obtain a reasonable model for the OF state GlpT.

6.2 Methods

6.2.1 System Preparation

For detailed description of modeling GlpT in the membrane, and simulation protocols refer to Section 3.2.2. 50 ns-equilibrated *apo* GlpT system and P_i^{2-}-bound GlpT system are used as the starting point of all simulations described in this chapter.

6.2.2 Simulation Systems

Simulation systems included in this section are 50 ns-equilibrated *apo* IF GlpT system, 50 ns-equilibrated P_i^{2-}-bound IF GlpT system. The substrate binding free energy simulation to the IF state were started from P_i^{2-}-bound IF GlpT system and is referred to as IF-system 1. Two different substrate binding free energy simulations were prepared for the OF state GlpT. These systems are referred to as OF-system 1, and OF-system 2 depending on the initial conformation of the system (see Section 6.2.5).

6.2.3 Exploration of IF↔OF Transition Pathway

The exploration of the IF↔OF Transition Pathway was performed by Mahmoud Moradi, PhD, using a novel approached elaborated in reference [140]. The approach is primarily based on the application of nonequilibrium driven MD simulations. The mechanistically relevant collective variable space is explored by various time-dependent biasing protocols.

The effectiveness of each protocol is assessed by the nonequilibrium work required to complete the transition. Numerous different driving protocols are initially explored in short simulations and those resulting consistently in lower nonequilibrium work are selected for subsequent longer simulations. The optimized biasing protocol involves two collective variables that describe orientations of two transmembrane helices. This protocol used in extended simulations starting from both *apo* and P_i^{2-}-bound GlpT systems.

6.2.4 Conformational Free Energy Calculations

Bias exchange umbrella sampling simulations were performed by Mahmoud Moradi, PhD. The centers and the force constants of the biases were optimized in short simulations to achieve ∼35% exchange rate between replicas. For free energy calculations of both *apo* and P_i^{2-}-bound GlpT, 12 replicas were used with initial conformations sampled from the extended nonequilibrium simulations described in Section 6.2.3 .

6.2.5 Substrate Binding Free Energy Calculations

Substrate binding free energy to IF-state GlpT was calculated using BEUS simulations starting with the 50 ns equilibrated P_i^{2-}-bound IF GlpT structure (IF-system 1). The reaction coordinate was chosen as the z-component of the distance between the phosphorus atom of P_i^{2-} and the center of mass C_α of R45, H165, and R269. Starting configurations were prepared by replacing the water molecule closest to the centers with a P_i^{2-}. Subsequently, 5000 steps of minimization and 0.1 ns of simulation were performed for each starting configuration with the protein heavy atoms restrained harmonically ($k = 10\,\text{kcal/mol/Å}^2$) and P_i^{2-} phosphorus atom restrained harmonically to the chosen center ($k = 1000\,\text{kcal/mol/Å}^2$). The centers and the force constants were optimized in several short simulations to optimize the exchange rate between the windows and achieve a full coverage of the reaction coordinate. Each of the 30 replicas was simulated 10 ns. Optimized centers are reported in Table 6.1, which also shows force constants and the calculated exchange rates for this BEUS simulation.

Two separate BEUS simulations for substrate binding to OF-state were performed

Table 6.1: The 30 windows used in BEUS simulations starting from the P_i^{2-}-bound IF-state GlpT and the exchange ratios between the neighboring windows.

center$_i$ (Å)	center$_j$ (Å)	k$_i$ (kcal/mol/Å^2)	k$_j$ (kcal/mol/Å^2)	accepted attempts	total attempts	ratio
-45	-42	0.002	0.002	4530	9999	0.45
-42	-35	0.002	0.002	2507	9999	0.25
-35	-30	0.002	0.002	1218	9999	0.12
-30	-26.5	0.050	0.050	2754	9999	0.28
-26.5	-25	0.050	0.050	4075	9999	0.41
-25	-24	0.050	0.050	4373	9999	0.44
-24	-23	0.050	0.050	1448	9999	0.14
-23	-22	1.000	1.000	3025	9999	0.30
-22	-21	0.500	0.500	2905	9999	0.29
-21	-20	0.500	0.500	2555	9999	0.26
-20	-19	0.500	0.500	1006	9999	0.10
-19	-18	0.500	0.500	3711	9999	0.37
-18	-17.5	1.000	1.000	4274	9999	0.43
-17.5	-17	2.000	2.000	3408	9999	0.34
-17	-16.5	3.000	3.000	2585	9999	0.26
-16.5	-16	4.000	4.000	2161	9999	0.22
-16	-15.5	3.000	3.000	2589	9999	0.26
-15.5	-15	2.000	2.000	3089	9999	0.31
-15	-14	1.000	1.000	1857	9999	0.19
-14	-13	1.000	1.000	1668	9999	0.17
-13	-12	1.000	1.000	1788	9999	0.18
-12	-11	1.000	1.000	1112	9999	0.11
-11	-10.5	1.000	1.000	2044	9999	0.20
-10.5	-10	1.000	1.000	2574	9999	0.26
-10	-9.5	1.000	1.000	3760	9999	0.38
-9.5	-9	1.000	1.000	4063	9999	0.41
-9	-8.5	1.000	1.000	4540	9999	0.45
-8.5	-8.3	1.500	1.500	3213	9999	0.32
-8.3	-8.15	3.000	3.000	3063	9999	0.31

each starting with an OF-state structure sampled from the conformational free energy simulations, one representing of the conformation free energy minimum of OF-state in the P_i^{2-}-bound GlpT (OF-system 1) and the other, in *apo* GlpT (OF-system 2). Starting configurations were prepared as described above for 30 replicas for each system. Each replica was run for ∼5 ns for the OF-system 1 and ∼15 ns for the OF-system 2.

6.2.6 Simulation Protocols

The simulations were performed with NAMD 2.6 [163] using the CHARMM27 force field with ϕ/ψ cross term map (CMAP) corrections [161]. All the production simulations were carried out using a 2-fs time-step, at a constant temperature of 310 K, and a constant pressure of 1 atm maintained only along the z direction, i.e., normal to the membrane (NP_nT). Water molecules were modeled as TIP3P [179]. The force field parameters for P_i^{2-} was adopted from similar molecules ($Me{-}PO_4^-$, $Me{-}PO_4^{2-}$, and glycerol) in the CHARMM force field. Constant pressure was maintained by the Nosé-Hoover Langevin piston method [180,181], and constant temperature by Langevin dynamics with a damping coefficient of 0.5 ps^{-1} for non-hydrogen atoms. Short-range interaction cutoff was set to 12 Å. Long-range electrostatic interactions were computed using the particle mesh Ewald (PME) method [182] with a grid density of at least 1 $Å^{-3}$. Bonded, non-bonded, and PME calculations were performed at 2-, 2-, and 4-fs intervals, respectively.

6.3 Results and Discussion

6.3.1 Characterization of the IF↔OF Transition

Intrigued by the relatively recently crystallized OF-state fucose transporter (FucP) [61] and know-how transferred to our group by Dr. Mahmoud Moradi, we attempted the modeling efforts once again. Dr. Mahmoud Moradi, through an extensive exploratory search of the phase space by non-equilibrium simulations, could distill potential transition pathways and OF-state models to the change of orientation in two transmembrane

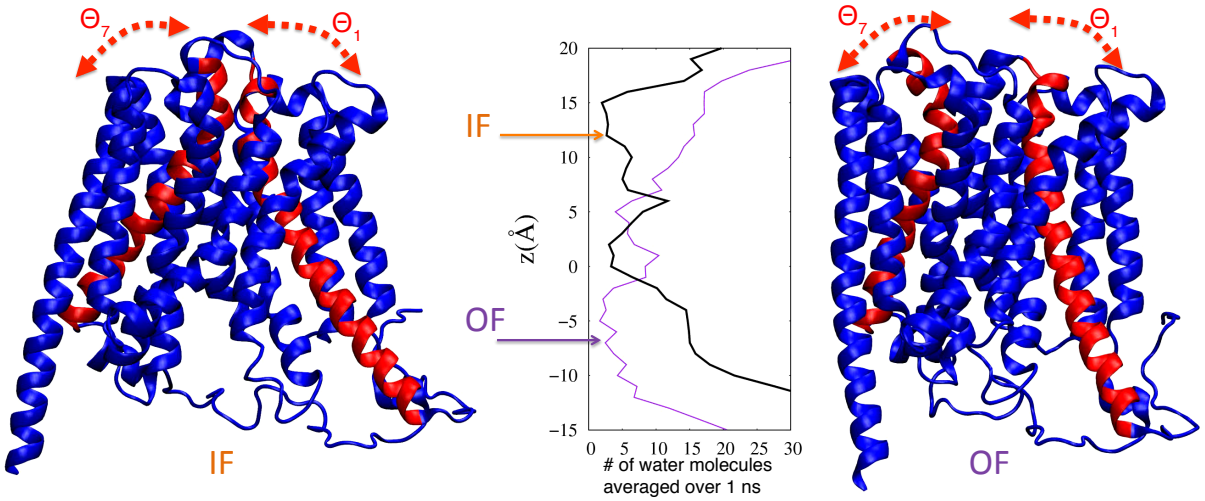

Figure 6.1: Modeling of OF state GlpT. Helices 1 and 7, whose orientations were changed in nonequilibrium simulations are shown in red (left) IF conformation of GlpT. (middle) The number of water molecules averaged over 1 ns simulation in IF simulation (black) and OF simulation (violet) are shown as a function of z-coordinate. IF state is closed on the cytoplasmic side and open on the periplasmic side. (right) Modeled OF is closed to cytoplasmic side and open to the periplasmic side. [a]

[a]Figure is a courtesy of Mahmoud Moradi, PhD.

helices (helices 1 and 7; colored red in Fig. 6.1). The exploratory search was performed by pulling and pushing different parts of GlpT along several reaction coordinates that could potentially materialize the proposed "rocker/switch mechanism". Several trials and protocols used in "pulling/pushing" could, then, be compared based on the resulting non-equilibrium work and attainment of a reasonable OF-state structure at the end of the simulation. This approach for probing conformational transitions have recently been applied to an ABC transporter [140]. Once a reasonable protocol for "pushing/pulling", which results in low nonequilibrium work, is obtained, extended nonequilibrium simulations were performed to investigate near quasi-static dynamics of the protein and optimize the characterized pathway further. Fig. 6.1 shows the starting IF state model, the helices (red) whose orientations were controlled during the transition, and the resulting OF-state model (right). Both the transition pathway and the resulting OF-state model are in agreement with the predictions of "alternating-access mechanism"/"rocker-switch model", resulting in the closing of the cytoplasmic side first (despite the helices whose orientation is controlled are not directly involved in the closing of the cytoplasmic side) and

opening of the periplasmic side. The OF-state model of GlpT is also supported by the OF state FucP structure (RMSD = 3Å) [61]. More interestingly, helices 1 and 7 contain the residues of the periplasmic salt bridges [165,170], which our simulations and several experiments predicted to be important in the IF↔OF transition and most of the contribution to the nonequilibrium work appeared to be from breaking the periplasmic salt bridges. Repeating exactly the same protocol, only starting with the P_i-bound GlpT [5, 165], instead of *apo* GlpT, on the other hand resulted in consistently less nonequilibrium work, indicating that the IF↔OF is, indeed, eased by substrate binding and that is due to the destabilization of the periplasmic salt bridges by the bound substrate.

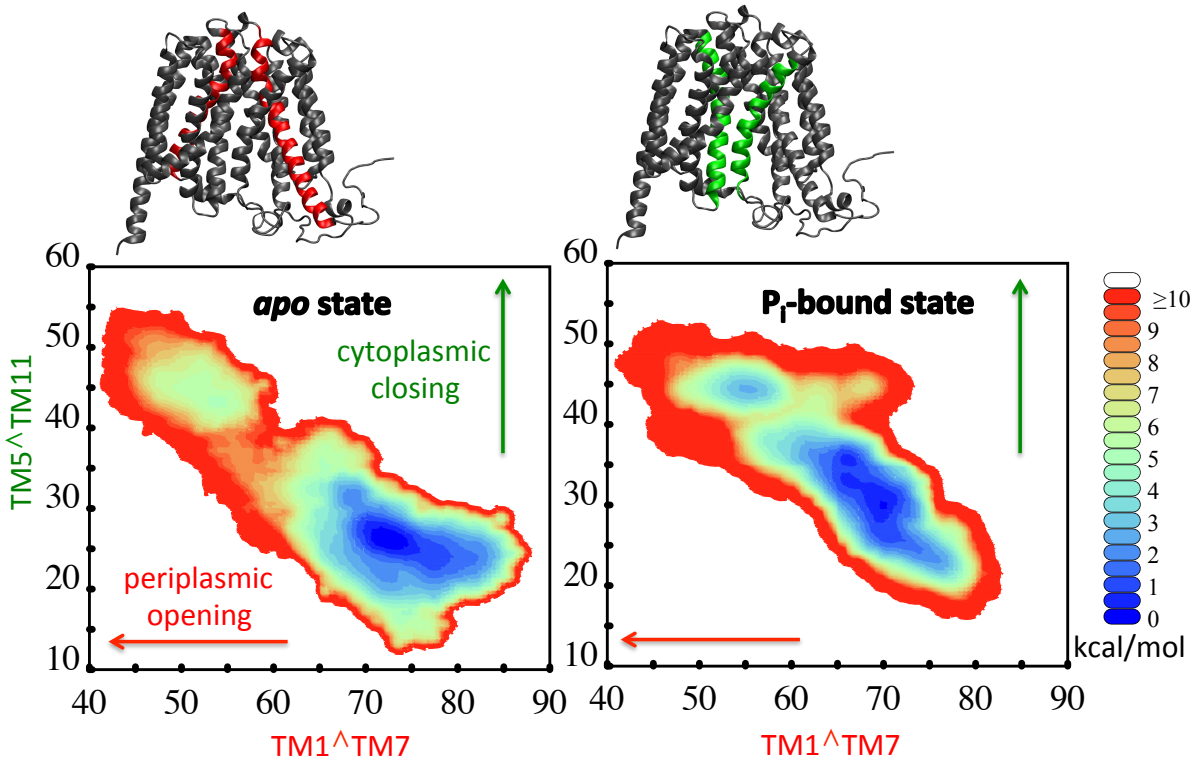

Figure 6.2: Two dimensional free energy plots shown calculated using BEUS for the IF↔OF transition in *apo* GlpT (left) and P_i-bound GlpT (right). Red helices are TM1 and TM7; green helices are TM5 and TM11. [a]

[a]Figure is a courtesy of Mahmoud Moradi, PhD.

To determine the equilibrium free energies associated with the identified IF↔OF transition, bias exchange umbrella sampling (BEUS) simulations [140,154–156] were performed by Dr. Mahmoud Moradi using the orientations of helices 1 and 7 as the reaction coordi-

nate. These simulations were then analyzed using GWHAM [140, 159, 160] and projected onto a two dimensional space that also involves the angles between helices 5 and 11 as the second dimension (Fig. 6.2), which we have previously identified as the main contributor to closing of the cytoplasmic side (Fig. 3.9). The 2D plots show that not only the OF state is stabilized further in the substrate bound case (Fig. 6.2, right), but also the barrier between them is reduced by ∼2.5 kcal/mol.

6.3.2 Substrate Binding Free Energies to IF and OF states

The availability of an OF-state model for GlpT made it possible to investigate the full transport cycle and associated free energies. We performed BEUS simulations to characterize the free energy of binding to IF- and OF-state GlpT. The first set of BEUS simulations (IF-system 1) were set up using the equilibrated P_i^{2-}-bound IF GlpT structure [165] to calculate PMF as a function of the z-component of the distance between the phosphorus atom of P_i^{2-} and the center of mass C_α of R45, H165, and R269 (Fig. 6.3, black). The binding free energy in this conformation is calculated to be >4 kcal/mol.

To characterize the free energies for the OF-state, we started with one of the structures representative of the minimum for the OF state in the conformational free energy map calculated for the P_i^{2-}-bound GlpT (OF-system 1), described before. After several rounds of optimization with 30 windows along the same reaction coordinate used for the IF-state, we came to realize that the structure is more likely a representative of an OF-occluded state, then an OF-state, due to the discontinuities in sampling and a ∼20 kcal/mol barrier along the reaction coordinate (Figure 6.3, green). To get a more reliable free energy estimate, we picked another structure from the *apo* GlpT free energy simulations that correspond to the minimum for the OF-state (OF-system 2). We, again, performed optimization, and 15 ns of BEUS simulations along the same reaction coordinate. The free energies reveal a binding free energy minimum that corresponds to a region closer to the periplasmic side than that found when started with the P_i^{2-}-bound GlpT, both in IF- and OF-states (Figure 6.3, red). These results along with the detailed examination of the structures obtained in the simulations indicate a significant rearrangement of the binding

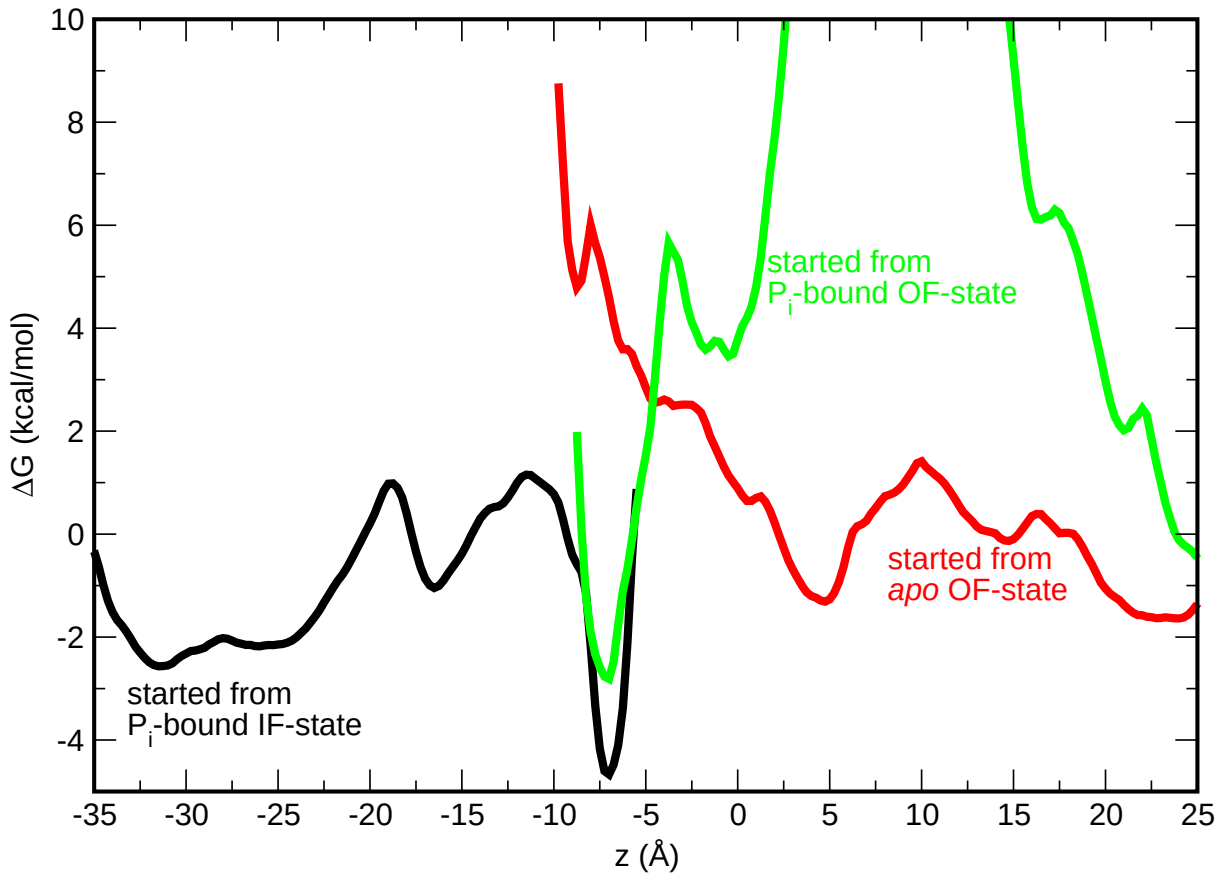

Figure 6.3: Free energies of substrate inside the GlpT lumen as a function of z-coordinate. The black line represents calculations started from the P_i^{2-}-bound IF, the red line represents those from the *apo* OF and the green line those from the P_i^{2-}-bound OF GlpT. Note that the values of the free energies cannot be compared directly between different plots.

site residues. In the *apo* OF-state, R45 and R249 extend towards the periplasmic side, differently from the crystal structure and other BEUS simulations, where they both extend to the intro-cellular side. These results indicate that there is a high free energy barrier for the reorganization of the binding site upon substrate binding and this may require an additional biased protocol that accounts for the rearrangement of the substrate binding site.

6.4 Summary

The exploration of the collective variable space using nonequilibrium simulations did not only result in characterization of the OF state of GlpT, but also allowed us to calcu-

late free energies associated with IF↔OF transition. We also calculated free energies of binding to both IF and OF states since substrate binding is one of the key events in the transport cycle. The current picture of the transport cycle of GlpT, we inferred from the conformational and substrate binding free energy calculations, show that substrate binding stabilizes the OF- and IF- state GlpT in relatively occluded conformations, while apo GlpT can sample more "open" conformations in both states. Based on the analysis of the structures and analysis, we could put forward a thermodynamic cycle (Fig. 6.4). Since our simulations has not been able to capture the rearrangement of the binding site spontaneously, elucidating the missing part of the cycle requires design of a new set of collective variables to describe the rearrangement and calculation of free energies associate with it. We believe that the rate-limiting step in the transport cycle of GlpT may, indeed, be the rearrangement of the substrate binding site.

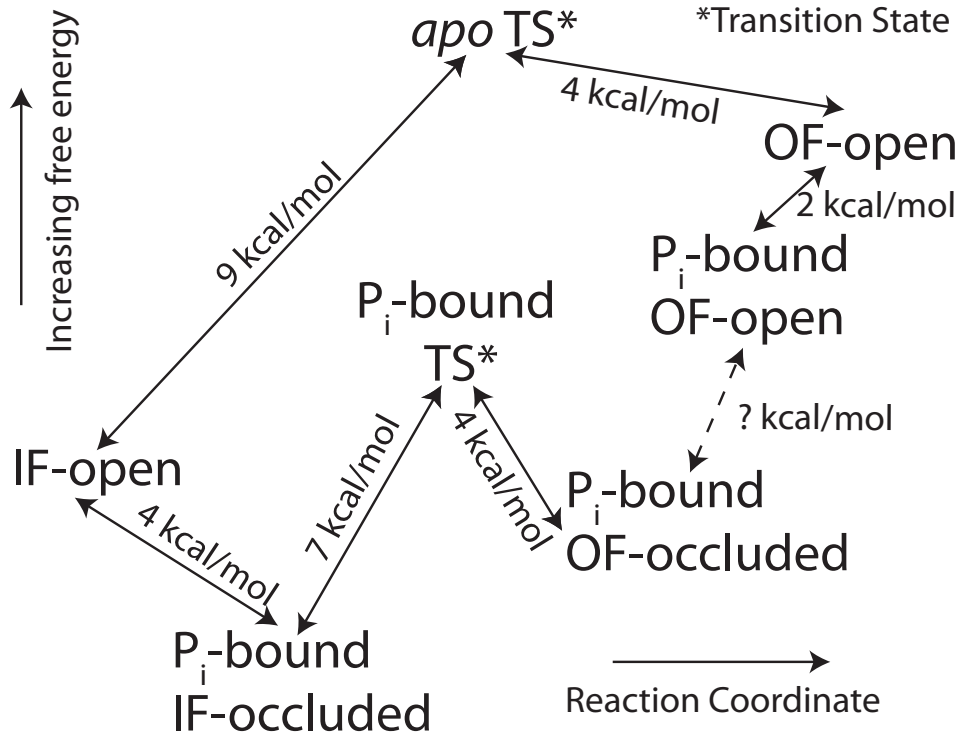

Figure 6.4: The IF↔OF transition free energies calculated from BEUS simulations. The free energies associated with the transition form P_i-bound OF-open state to P_i-bound OF-occluded state are unknown.

Chapter 7

Structure and Permeation Mechanism of Urea Transporter *

7.1 Introduction

Urea transporters (UTs) are integral membrane proteins that facilitate diffusion of urea down its concentration gradient. UTs were originally predicted to be transporters. However, measurements of single-channel flux rates ranging from 10^4–10^6 urea molecules/s [233, 234] were more consistent with a channel-like mechanism. This was confirmed with the solution of the structure of a bacterial homolog [83], dvUT, which forms a trimer with a continuous membrane-spanning pore at the center of each protomer. The structures of dvUT were either in their substrate-free (*apo*) form, or a substrate analog, dimethylurea (DMU), Due to its bulkier size compared to the natural substrate (urea), DMU cannot pass through and bound to two sites (S_i and S_o), respectively located at the internal (S_i) and external (S_o) openings of the channel (Fig. 7.1) [83]. Two structurally-symmetrical threonines in optimal positions for hydrogen bonding along with the formation of a small cavity suggested a middle binding site (S_m) and a putative urea binding mode [83,233,235].

This chapter appeared as a research article in Proceedings of the National Academy of Sciences [49]. Elena J. Levin, Yu Cao, Giray Enkavi, Matthias Quick, Yaping Pan, Emad Tajkhorshid, Ming Zhou*, "Structure and permeation mechanism of a mammalian urea transporter". *Proceedings of the National Academy of Sciences*, 109(28):11194–11199, **2012**.

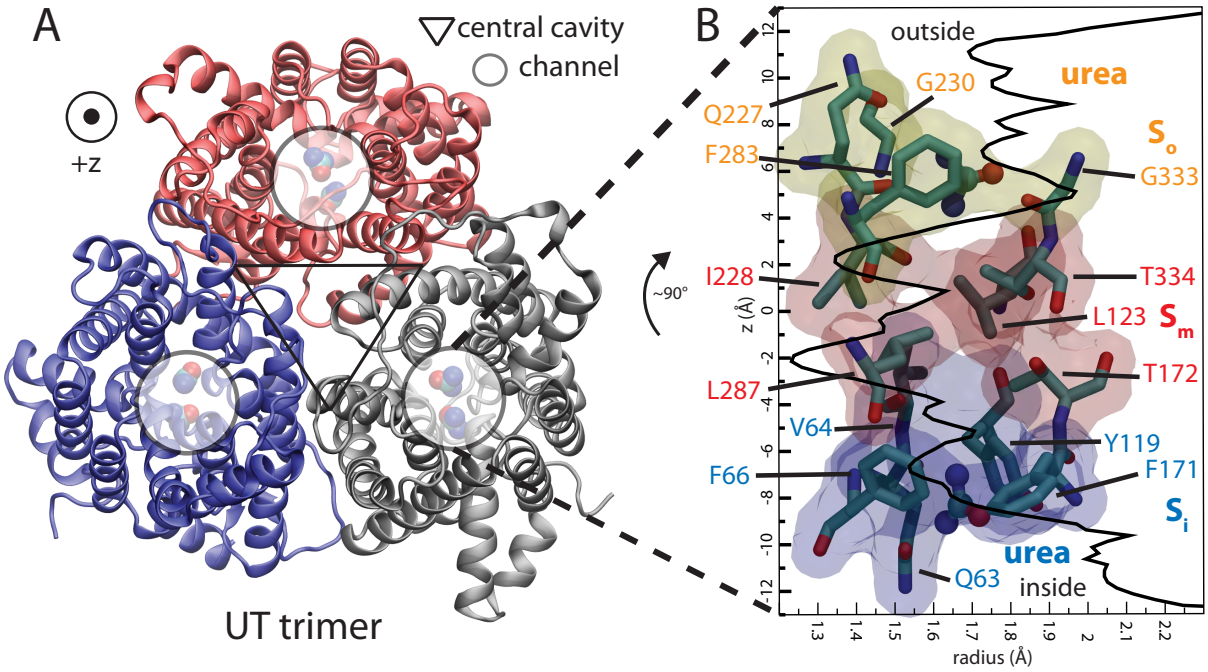

Figure 7.1: Structure of UT-B. (A) Trimeric Structure of UT-B shown from the extracellular side. Individual monomers are colored differently. Monomeric channels are indicated with a circle, while the central cavity is indicated with a triangle. (B) The residues that line the monomeric channels. Putative urea binding sites are labeled as S_i, S_m, and S_o overlaid with the radius of the channel as a function of the channel axis.

The role of the putative S_m site in the transport and how urea penetrates into this site are unknown. Likewise, investigation of how these binding sites (S_i, S_o, and S_m) confer selectivity for urea also requires a dynamical and atomic resolution view. Furthermore, due to the high polarity of urea, water permeability of UT is closely associated with urea permeation and has possible implications for urea transport [83]. Permeation of highly hydrophilic urea may require coordination by water, or the channel could catalyze dehydration of urea as it goes through [83]. Besides, it remained unclear how similar dvUT structure was to that of the mammalian UTs, and the details of the permeation mechanism were unknown. To answer these questions, our collaborators, Ming Zhou lab (Columbia University, NY)[†] solved the structure of a mammalian UT-B and performed functional studies. We performed molecular dynamics simulations to characterize energetics of urea permeation through the monomeric pores.

[†]Current position at Baylor College of Medicine, Houston, TX

7.2 Methods

7.2.1 System Preparation

The selenourea/UT-B trimer was used as the initial structure (PDB ID 4ezd). The two bound selenourea molecules were replaced by urea, and the three lipid molecules in the central cavity of the trimer were modeled as three POPC molecules that were subsequently minimized to remove steric clashes. After removing all other detergent molecules, the trimer was embedded into a POPC lipid bilayer ($120 \times 120\,\text{Å}^2$), with the membrane normal aligned along the z axis. After removing the lipids overlapping with the protein trimer, the system was solvated and ionized with 100 mM NaCl by randomly replacing water molecules with Na^+ or Cl^- ions, resulting in system dimensions of $120 \times 120 \times 90\,\text{Å}^3$ and approximately 109,000 atoms.

The lipid tails were then "melted" using 5000 steps of energy minimization and 1 ns of constant volume and temperature (NVT) MD simulation at 310 K while all other atoms of the system were fixed, in order to allow the initially ordered lipid tails to maximize their disorder and partially pack against the protein. In the next step, the whole system was simulated for 2 ns under constant pressure and temperature (NPT) conditions while all heavy atoms of the protein and the urea molecules were harmonically restrained ($k = 2\,\text{kcal/mol/Å}^2$) to allow further relaxation and packing of the lipids against the protein. During this phase, water molecules were prevented from entering the hydrophobic core of the membrane by employing additional constraints. The resulting relaxed configuration of the urea/UT-B system was used as the starting structure for the production simulations described below.

7.2.2 Simulation Systems

The apo-UT-B system was generated by removing the urea molecules from the system. T172S/T334S and T172V/T334V double mutant systems were constructed from the apo-UT-B system by mutating the respective residues and minimized for additional 10,000 steps before subjected to MD simulations. The four simulation systems, namely,

urea/UT-B, apo-UT-B, T172S/T334S, and T172V/T334V, were then each simulated for approximately 60 ns, 100 ns, 45 ns, and 25 ns, respectively, under isothermal-isobaric NPT conditions.

7.2.3 Umbrella Sampling Simulations of Urea

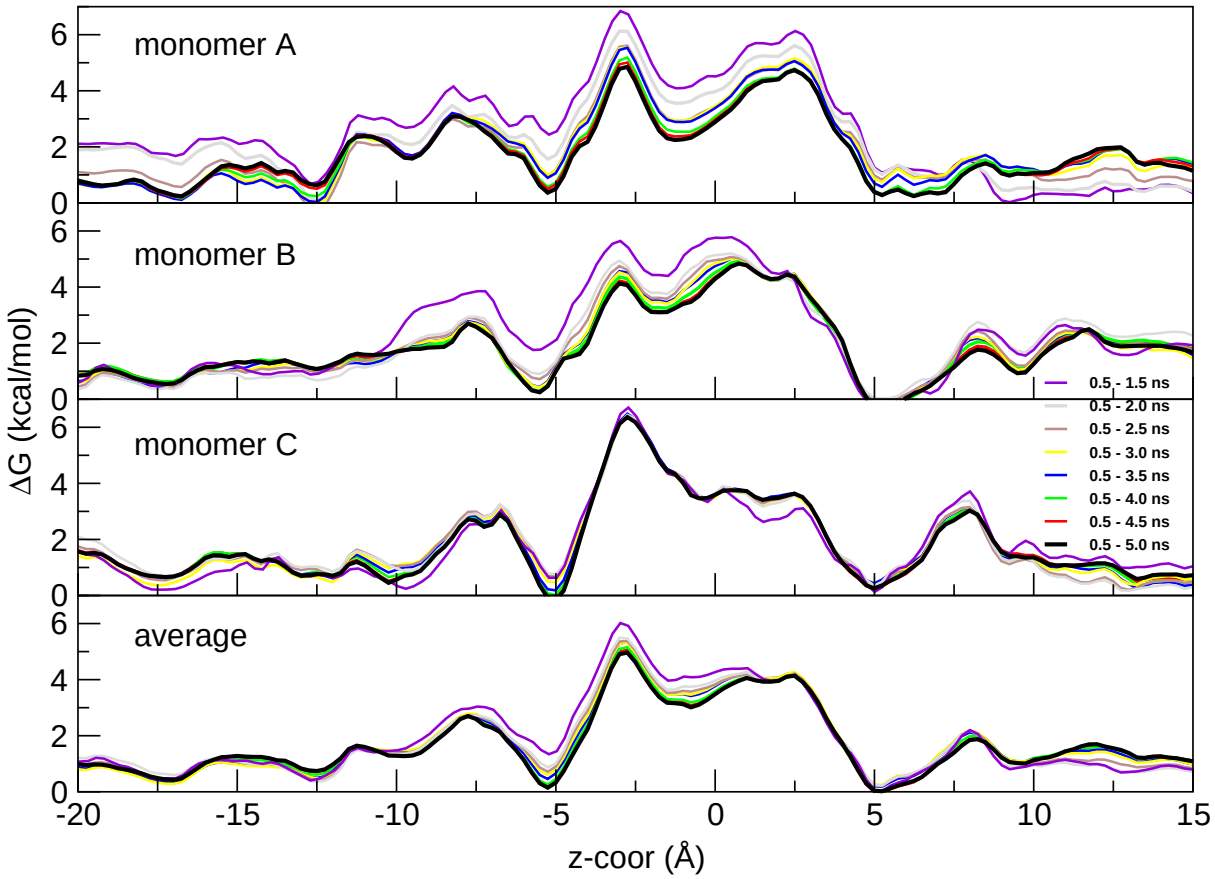

Figure 7.2: Convergence of potential of mean force. PMFs for urea permeation calculated using sequentially increasing intervals from the umbrella sampling simulations. The changes in the PMFs are minimal after approximately 3 ns, indicating convergence of the simulations.

In order to reconstruct the potential of mean force (PMF), umbrella sampling (US) simulations were initiated using the 50-ns equilibrated structure from the apo-UT-B simulation. 71 umbrella windows of 0.5 Å each were defined along the channel axis, covering a range from z = -20 — +15 Å with the origin (z = 0) at the center of mass (COM) of the C_α atoms of channel-lining residues (S_i, S_m, and S_o) of each monomer. The starting configuration for each US simulation was generated by adding a urea molecule to each monomer

with the position of its carbon atom harmonically- ($k = 10\,\text{kcal/mol/Å}^2$) restrained to the center of the respective window only along the z axis. Each starting configuration was separately minimized for 5000 steps and simulated for 5 ns. The restraining potential and the position of the substrate for each monomer were recorded at 0.1-ps intervals. Including only the last 4.5 ns of the US simulations (45,000 data points for each monomer and each window), the weighted histogram analysis method (WHAM) [157, 236] was used to reconstruct the PMF with 0.25 Å resolution for each monomer along with a combined PMF. The quality of the PMF was tested by examining the resulting profile from shorter simulation times (checked for 0.5 ns increments) for each window, clearly indicating convergence(7.2).

7.2.4 Simulation Protocols

All the simulations were performed using NAMD 2.8 [163] with CHARMM27 force field with ϕ/ψ cross term map (CMAP) corrections [161] and CHARMM36 all-atom additive parameters for lipids [162]. Force field parameters for urea were adopted from Caflisch *et al.* [237], and water was modeled as TIP3P [179]. All production simulations were maintained at 1.0 atm pressure using the Nosé-Hoover Langevin piston method [180, 181] and at 310 K temperature using Langevin dynamics with a damping coefficient of $0.5\,\text{ps}^-1$ applied to all non-hydrogen atoms. Short-range interactions were cut off at 12 Å with a smoothing function applied after 10 Å, and long-range electrostatic forces were calculated using the particle mesh Ewald (PME) method [182] at a grid density of $>1\,\text{Å}^{-3}$. Bonded, non-bonded, and PME calculations were performed at 2-, 2-, and 4-fs intervals, respectively.

7.2.5 Summary of Experimental Methods

Our collaborators, Ming Zhou lab (Columbia University, NY) [‡], performed cloning, expression and purification of wild type and mutant UT-B, solute flux assays and oocyte uptake assays. They, also, crystallized the *apo* and selano-urea bound structures [49].

[‡]Current position at Baylor College of Medicine, Houston, TX

7.3 Results and Discussion

7.3.1 Crystal Structure of Bovine UT-B

The structure was solved to a resolution of 2.36 Å by molecular replacement using the structure of dvUT [83] (PDB ID 3k3f) as a search model. Final model contains three UT-B protomers with residues 31 to 376 resolved and 23 complete and partially ordered detergent and lipid molecules in the asymmetric unit.

Bovine UT-B forms a trimer (Fig. 7.3) with a total buried surface area of approximately 3500Å^2. Furthermore, the trimer interface is formed by equivalent helices in both the dvUT and UT-B structures, indicating that this quaternary structure may be conserved across the UT family. At the center of the trimer interface is a large cavity sealed off from the solvent, which is packed with partially ordered lipid or deter- gent molecules.The individual protomers have the same overall fold as dvUT, and the root mean square deviation for main chain atoms in the transmembrane region is 0.7 Å. The UT fold contains two homologous halves with opposite orientations in the membrane, likely the product of duplication of an ancestral gene [238, 239], which give the structure an internal pseudo-twofold symmetry axis. Each half contains five transmembrane helices (T1a-5a and T1b-5b, Fig. 7.3D) and one tilted reentrant helix spanning roughly half of the membrane (Pa and Pb, Fig. 7.3B and D). An amphipathic membrane-spanning pore is formed at the interface of the two halves in each protomer and is lined by residues from conserved urea signature sequences [238]. The selectivity filter can be divided into three regions: the S_o and S_i sites, which are rectangular in cross-section with evenly-spaced arrays of carbonyl and side-chain oxygen atoms lining one side of the pathway (the oxygen ladders); and S_m, a constricted region at the center of the filter which is hydrophobic save for a pair of pseudo-symmetry related threonine residues (Figs. 7.3C, and 7.4A–B).

7.3.2 Ligand Binding Sites in Selectivity Filter

Then, Zhou lab attempted to solve a substrate-bound structure to visualize interactions between the substrate and the protein. Since efforts to obtain high-resolution crystals of

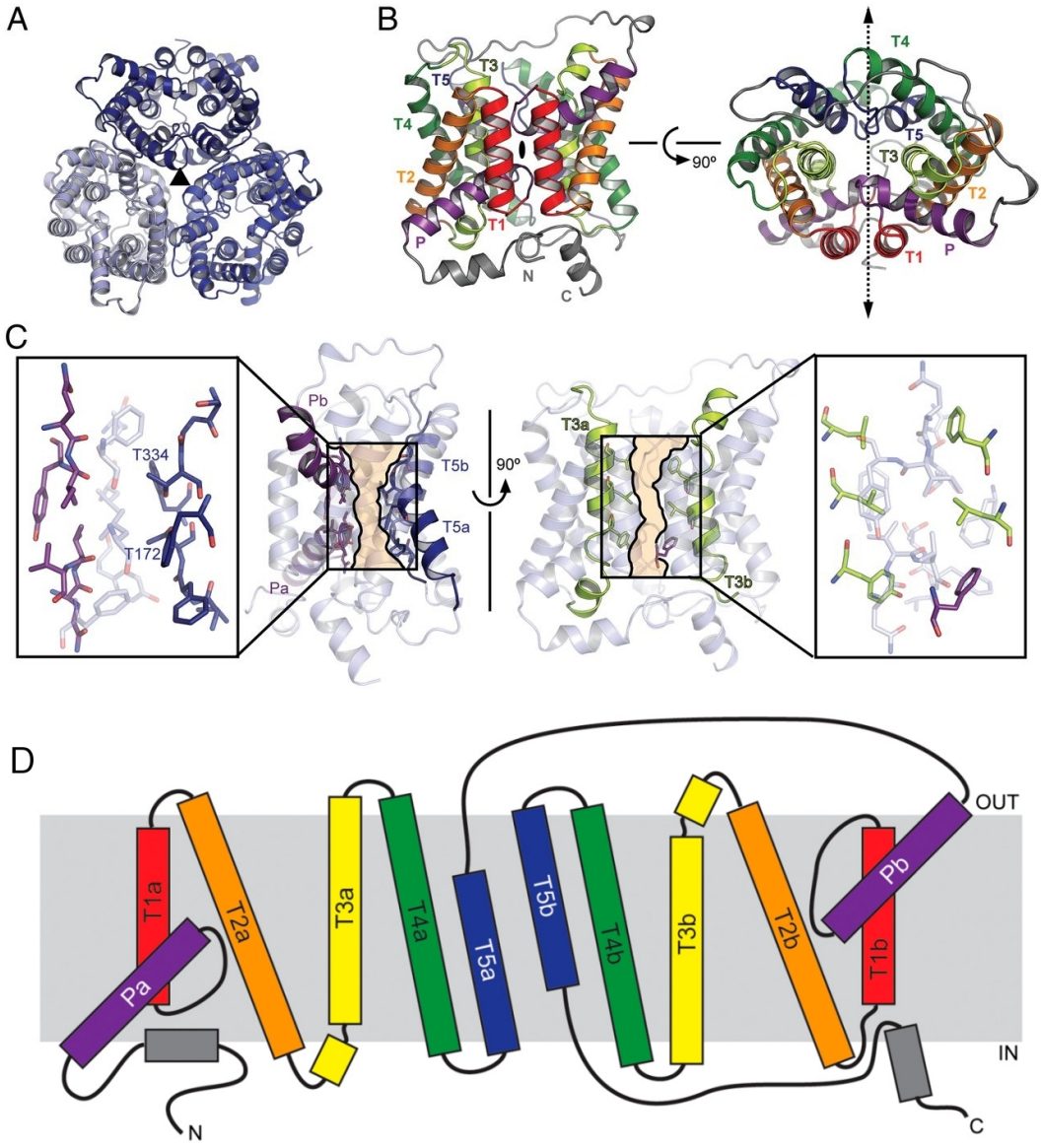

Figure 7.3: 2.36 Å crystal structure of bovine UT-B. (A) The UT-B trimer viewed from the extracellular side of the membrane, with the threefold symmetry axis (black triangle) labeled. (B) The structure of a UT-B protomer as viewed from within the plane of the membrane (Left) and from the extracellular side of the membrane (Right). The transmembrane helices are colored in pseudo-symmetry-related pairs. (C) The pathway of urea permeation is shown viewed from two perpendicular orientations, with cross-sections of the wide (Left) and narrow (Right) dimensions of the pore marked in beige. Insets show residues lining the pore within the regions marked by black rectangles. (D) Diagram of the shared membrane topology of dvUT and UT-B, oriented with the intracellular side on bottom. Pairs of pseudo-symmetric, equivalent helices between the two homologous repeats are shown with the same color; dark gray helices are not involved in the homologous repeats. [a]

[a]This figure was prepared by the other co-authors of the research article in Proceedings of the National Academy of Sciences [49].

the protein bound to urea were unsuccessful, UT-B was cocrystallized with selenourea, which is chemically similar to urea and has the additional advantage of producing a strong anomalous signal. The structure of UT-B in complex with selenourea was solved by molecular replacement to a resolution of 2.5 Å.

The protein's conformation is unchanged, but two wedge-shaped densities consistent with selenourea appear in the S_o and S_i sites (Fig. 7.4), coincident with two >13.0 anomalous difference peaks that render the identity and orientation of the selenourea molecules unambiguous (Fig.7.4B). Both substrates are oriented to form hydrogen bonds between their amide hydrogens and the oxygen ladders. However, the two sites are not pseudo symmetrically equivalent: The S_o site selenourea is positioned to straddle the outermost "rung" of the oxygen ladder, while the S_i site ligand is level with the middle carbonyl. This suggests that as urea moves through the S_o and S_i regions, it may have more than one stable position. The substitution of one of the phenylalanines in the S_o site, F190 in the bacterial structure, with a leucine residue at position 127 in the mammalian structure does not appear to have affected the suitability of this region as binding site, implying that aromatic residues are not specifically necessary for stabilizing the bound substrate. Intriguingly, the S_m site, which appears to have a suitable configuration of hydrogen bond donors and acceptors for urea binding, was unoccupied in the selenourea-bound UT-B structure as well as in a previous dvUT structure bound to the urea analog 1,3-dimethylurea [83].

In order to provide a more dynamic description of the mechanism and pathway of permeation with the natural substrate, urea, and to characterize the energetics associated with the process, we then employed molecular dynamics simulations. Umbrella sampling was used to reconstruct the potential of mean force (PMF) associated with urea permeation through UT-B using the projection of the carbon atom of urea along the pore axis as the reaction coordinate. Urea permeates the channel with minimal rotation around the channel axis, a configuration imposed by the hydrogen bonding pattern with the oxygen ladder in an otherwise hydrophobic pore. The projection of the dipole moment of urea on the pore axis exhibits a clear pattern of alternation reflecting its stepwise transition be-

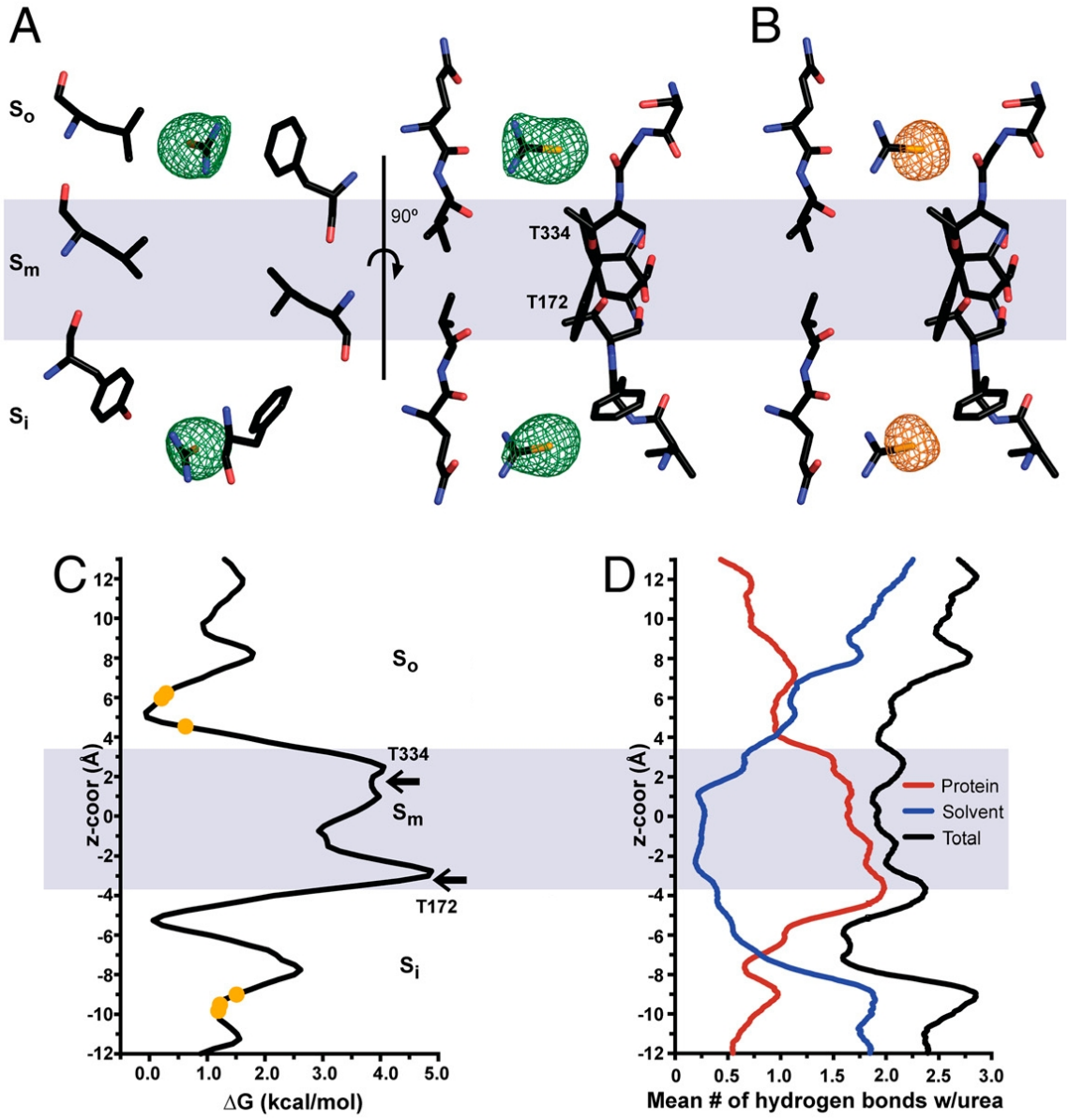

Figure 7.4: Energetics and binding sites in the UT-B pore. (A–B) [a] Electron density calculated without ligands in the model (A), and anomalous difference density (B) are shown in the pore of the selenourea/UT-B structure along with key selectivity filter residues. The maps are contoured at 3 and 5, respectively. (C) MD simulations were used to reconstruct the potential of mean force of urea permeation through the UT-B pore. The yellow circles indicate the z-coordinates of the central carbons of bound selenourea in the three subunits of the crystal structure, black arrows indicate the z-coordinates of the S_m site threonine C_γ atoms averaged over the three subunits. (D) The number of hydrogen bonds between urea and the protein (red), water molecules (blue), or both (black) averaged over the MD simulations as a function of the reaction coordinate.

[a]The upper panel of this figure was prepared by the other co-authors of the research article in Proceedings of the National Academy of Sciences [49].

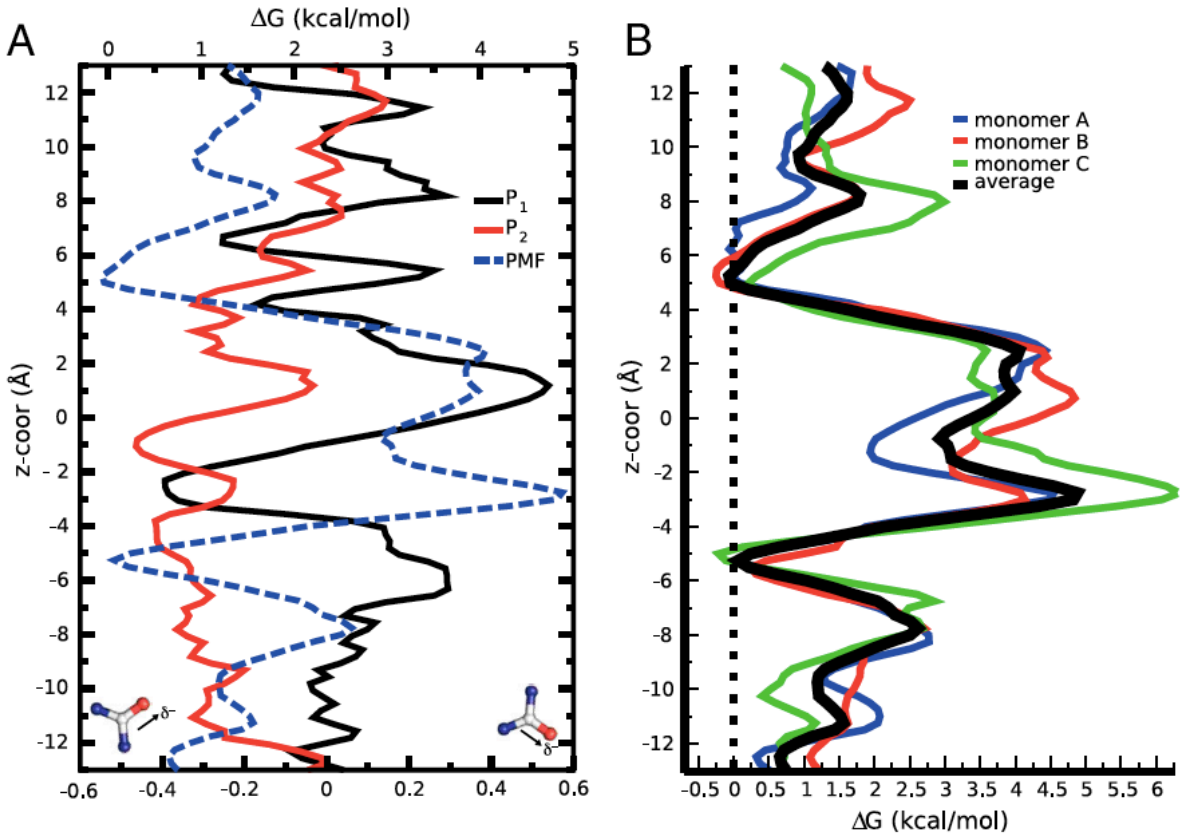

Figure 7.5: Substrate orientation and potential of mean force for urea permeation through UT-B. (A) Orientational order parameters (P_1 and P_2) of urea as a function of channel axis overlaid with the PMF calculated for urea permeation through UT-B. The polar order parameter, $P_1 =< \hat{d} \cdot \hat{z} >$, where $\hat{d}$ and $\hat{z}$ are the unit vectors of the dipole moment of urea and the channel axis, respectively, captures the orientation of urea with respect to the channel axis; the images of urea molecules along the Bottom of the figure indicate the direction of negative vs. positive values when the pore is oriented with the extracellular side on the Top. The nematic order parameter, $P_2 = \frac{3<(\hat{d}\cdot\hat{z})^2>-1}{2}$, is used to distinguish between isotropic and ordered states. (B) The PMF as a function of the channel axis for the three individual subunits in UT-B (red, blue and green) and the average over all three subunits (black).

tween adjacent, hydrogen-bonding sites (Fig. 7.5). This "rocking" motion allows urea to permeate through the channel without causing significant perturbation, confirming that UTs indeed operate by a channel-like mechanism.

The PMF shows two almost symmetric pairs of energy minima in the S_o and S_i regions (Fig. 7.4C, Fig. 7.5B). Two of these minima, one in each of the S_o and S_i sites, agree well with the observed ligand positions in the selenourea-bound structure (Fig. 7.4C, yellow circles). In contrast, the S_m site is characterized by a large energy barrier with a maximum G of approximately 5.0 kcal/mol with respect to the S_i or S_o sites. We attribute the energy

barrier at the S_m site primarily to the desolvation cost in this region; while in the S_o and S_i sites, urea is only partially dehydrated and maintains on the average approximately 1.5–2.0 hydrogen bonds with water molecules (Fig. 7.4D), whereas upon entering the S_m site it becomes completely dehydrated.

7.3.3 Role of Binding Sites in Permeation and Modulation

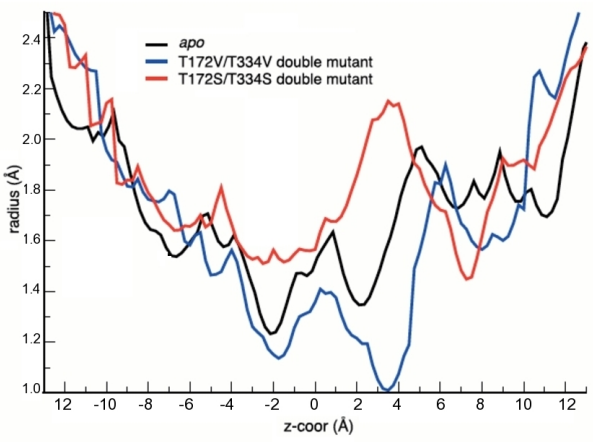

Figure 7.6: Radius of the UT-B pore. Radius of the pore along the channel axis in WT (black), T172V/T334V (blue), and T172S/T334S (red) bovine UT-B in the equilibrium simulations, calculated by the program HOLE [185].

To verify this observation experimentally, our collaborators at Zhou lab mutated an S_m site threonine to valine to further diminish the ability of the protein to compensate for the dehydration penalty of urea in that region. While the wild-type channel was functional, the rate of urea efflux for the T334V mutant was close to that of empty liposomes, as predicted by the simulations. In contrast, the T172S and T334S mutations, which conserve hydrogen bonding, had little effect on the rate of urea flux even when combined. Equilibrium MD simulations of both the T172V/T334V and the T172S/T334S double mutants show that in addition to increasing the hydrophobicity of the S_m site, loss of the hydrogen bond between T172 and T334 in the valine mutants causes a rotation of V334 towards the pore, so that the hydrophobic side chain obstructs the S_m site to the degree that even water molecules are not able to readily cross this region (Fig. 7.6).

These results indicate that the UT selectivity filter does not form a contiguous series

of binding sites, but instead contains a substantial energy barrier at the highly conserved S_m site. We speculated that relatively small changes in the conformation of the pore at the S_m site could have substantial effects on the rate of transport, and that S_m could potentially serve as a site to regulate the rate of urea permeation. Given that the physiological role of UTs is closely linked to osmoregulation, Zhou lab tested the channel for sensitivity to osmotic stress by measuring radio-labeled urea uptake into oocytes expressing UT-B in isotonic, hypotonic, and hypertonic buffer. Oocytes expressing wild type UT-B experienced an approximately twofold increase in urea uptake in the hypotonic buffer relative to uptake under isotonic conditions, indicating sensitivity of the channel to osmotic stress. In contrast, the hypertonic buffer did not alter the rate of urea uptake. This suggests that the effect cannot be attributed solely to cotransport of urea with water, since in that case reversing the direction of water permeation would be expected to have an inhibitory effect. To test whether the S_m site was involved in the response of the channel to hypoosmotic stress, urea uptake for oocytes expressing the T172S/T334S double mutant under identical conditions was also measured. Consistent with the results of the liposome assay, the T172S/T334S mutant had a similar basal rate of urea permeation to the wild type channel under isotonic conditions. However, unlike the wild type channel, the T172S/T334S mutant did not exhibit an increase in the rate of transport in a hypotonic buffer. These results demonstrate a potential mechanism for modulation of UT-B function and support the hypothesis that the rate of urea conduction is regulated at the constricted region, the S_m site, in the selectivity filter.

7.4 Summary

UT-B is a trimeric urea channel whose rate of permeation has been determined by Zhou lab to be modulated by osmotic pressure. The energy barrier at S_m in UT-B is reminiscent of that observed in the aromatic/arginine motifs (selectivity filter; SF) of aquaporins [112, 119, 120] or the conserved pore histidines in ammonia channels [240, 241]. Aquaporin SF are known to be crucial for selectivity; the height of this barrier largely

determines the difference between the canonical aquaporins that permeate only water and the comparatively promiscuous aquaglyceroporins [40, 110, 118]. The S_m site may likely play an analogous role in UTs, and confer the ability to block the flux of charged species like protons, ammonium and guanidinium that has been measured previously for other UT homologs [235, 242, 243]. While the S_m site does not contain charged residues to exclude ions through electrostatic repulsion, as the SF of aquaporins do [129], it is relatively hydrophobic and too narrow to permit a hydration sphere; the helix dipoles may also help repel anions. The hydrophobicity of the S_m site likely accounts for the lower rate of water conduction through urea transporters relative to the canonical aquaporins [244], despite the strong structural similarities of the pores between the two families.

One notable difference between the UT-B and UT-A orthologs is that the latter is upregulated by the antidiuretic hormone vasopressin via phosphorylation of multiple sites on its long cytoplasmic N-terminus [101]. Some of this increase in activity can be accounted for by increased localization of UT-A in the plasma membrane [245], but there is also evidence for an increase in urea transport activity that occurs on a more rapid time scale than the rate of accumulation of UT-A in the plasma membrane, possibly due to modulation of UT activity through phosphorylation of proteins already present at the cell surface [246]. Given the high conservation of pore-lining residues between UT-B and UT-A, it is tempting to consider the possibility that the Sm site barrier is also present in UT-A, but with phosphorylation rather than osmotic stress as the trigger for modulation.

In the kidney, UT-B is expressed in the descending vasa recta, the blood vessel that supplies the nephron. In addition to providing the nephron with blood, the vasa recta is thought to play a role in the counter-current exchange mechanism that prevents washout of the urea gradient in the inner medullary interstitium [247–249]. Mice deficient in UT-B experience a 40% urine concentrating deficiency, and a substantially lowered ability to retain urea in the inner medulla [91]. In contrast to UT-As in the inner medullary collecting ducts, which transport urea from the renal tubules into the hyperosmotic interstitium, UT-B mediates net efflux of urea out of the interstitium into the lumen of the vasa recta. Since the descending vasa recta also expresses aquaporins and is therefore highly water

permeable [250], it is conceivable that a mismatch in osmolarity between the vasa recta and the interstitium could trigger membrane stretch and thereby upregulate UT-B. The modulation of UT-B by osmotic stress observed in this study therefore appears consistent with the known physiological role of UT-B in the kidney. Further studies would be necessary, however, to know how significant such an effect would be, or to determine the structural changes underlying this modulation.

Chapter 8

Urea Transporter as a Gas Channel *

8.1 Introduction

Gases had been thought to cross biological membranes simply by dissolving in and then diffusing through the lipid phase of the membrane. Since the discovery that aquaporin 1 (AQP1) and Rh-associated glycoprotein (RhAG), which acts as ammonia channel, proteins highly expressed in the membrane of red blood cells, are capable of transporting both carbon dioxide (CO_2) [252–257] and ammonia (NH_3) [256, 258–260], the field of gas channels has been evolving rapidly [261]. Together, AQP1 and RhAG account for 90% of the CO_2 traffic across human red blood cell membranes [253, 254]. AQP1 can also transport another gas relevant to red blood cells, namely, nitric oxide (NO) [262, 263]. Based on these observations, it is of interest to determine if the urea transporter B (UT-B), a significant membrane protein in red blood cells, might also function as a gas channel and transport CO_2 or NH_3. Presumably, gas-channel specificities rely on four major criteria: the chemical and physical properties of the gas, and the chemical and physical properties of the pathway(s) formed by the protein. At the outset of this study, we hypothesized that water, CO_2, and NH_3 could pass through the hydrophilic, monomeric urea channel

*This chapter appeared as a research article in American Journal of Physiology – Renal Physiology [251]. R. Ryan Geyer, Raif Musa-Aziz, Giray Enkavi, Paween Mahinthichaichan, Emad Tajkhorshid, and Walter F. Boron, "Movement of NH_3 through the human urea transporter b: a new gas channel". *American Journal of Physiology – Renal Physiology*, 304(12):F1447–F1457, **2013**

of UT-B.

8.2 Methods

8.2.1 System Preparation

The two bound selenourea molecules in the selenourea/UT-B trimer (PDB ID 4ezd) [49] were replaced by urea and three POPC molecules were modeled into the central cavity. The trimer was, then, placed into a POPC lipid bilayer and solvated with 100 mM NaCl The lipid tails were then randomized in a 1 ns NVT simulation at 310 K. Then, the systems was simulated under NPT for 2 ns with protein heavy atoms restrained. The details of system preparation was previously explained in Section 7.2.1. The resulting configuration of the urea/UT-B system was used as the starting structure for the other simulation systems.

8.2.2 Simulation Systems

The apo-UT-B was prepared by removing the urea molecules from the urea/UT-B as described in 7.2.1 and simulated for 100 ns. The T172S/T334S and T172V/T334V double mutant systems and T172V and T334V single mutant systems were constructed from the apo-UT-B system by mutating the respective residues. The double mutant systems were minimized for 10,000 steps and were simulated for 45 ns and 25 ns as described in 7.2.1. The single-mutant systems were subjected to minimization for 1,000 steps followed by 0.3 ns of simulation in which all the protein heavy atoms except those of the mutated residues were harmonically restrained ($k = 5$ kcal/mol/Å^2). The single-mutant systems were later simulated for 10 ns each under NPT conditions.

8.2.3 Simulation Protocols

All the simulations were performed using NAMD 2.9 [163] with CHARMM27 force field with ϕ/ψ cross term map (CMAP) corrections [161] and CHARMM36 all-atom additive

parameters for lipids [162]. Force field parameters for urea were adopted from Caflisch *et al.* [237], and water was modeled as TIP3P [179]. All production simulations were maintained at 1.0 atm pressure using the Nosé-Hoover Langevin piston method [180, 181] and at 310 K temperature using Langevin dynamics with a damping coefficient of 0.5 $ps^{-}1$ applied to all non-hydrogen atoms. Short-range interactions were cut off at 12 Å with a smoothing function applied after 10 Å, and long-range electrostatic forces were calculated using the particle mesh Ewald (PME) method [182] at a grid density of >1 $Å^{-3}$. Bonded, non-bonded, and PME calculations were performed at 2-, 2-, and 4-fs intervals, respectively.

8.2.4 Umbrella Sampling Simulations of Ammonia

The PMF profile for ammonia (NH_3) was calculated using umbrella sampling (US), initiated from the 50-ns equilibrated structure of the apo-bUT-B simulation. In each US set, 53 umbrella windows at 0.5- intervals were defined along the channel axis, covering a range from z = -13 — +13 Å with the origin (z = 0) at the COM of the C_α atoms of channel-lining residues (from the S_i, S_m, and S_o sites) of each monomer. The water molecule closest to the center of each window in each monomeric pore was replaced by the ligand (NH_3) to generate the starting configurations. Each window was initially minimized for 1,000 steps using the conjugated gradient algorithm. Then, a 2-ns US simulation was carried out with the position of the nitrogen atom of NH_3 being restrained only along the z-axis by a harmonic potential, $Ui = \frac{k}{2} \times (z - z_i)^2$, where k = 3 kcal/mol/$Å^2$ and z_i is the center of the respective window. Half-harmonic restraints, in the form of $U_c = \frac{k_c}{2} \times \max[0, (x - x_0)^2 + (y - y_0)^2 - r2]$, where k_c = 10 kcal/mol/$Å^2$, r = 5 Å and x_0 and y_0 are the x and y components of COM of the C_α atoms of the residues of the S_i, S_m, and S_o sites of each monomeric pore and were also applied to confine the sampling to a cylindrical region encompassing each monomeric pore. The z-coordinates of the nitrogen atom of NH_3 in each monomeric pore were recorded at 0.1-ps intervals. Including only the last 1.5 ns of the US simulations (15,000 data points for each monomer, each window, and each set), the weighted histogram analysis method (WHAM) [157, 236] was

used to reconstruct the PMF profiles at 0.25 Å bin width, which were later normalized with respect to the average PMF at S_i and S_o sites.

8.2.5 Specific Analysis

Histograms of z-coordinate positions of water within the cylindrical region of $(x - x_0)^2 + (y - y_0)^2 < r^2$, where $r = 10$ Å and x_0 and y_0 are the x and y components of the center of mass (COM) of the C_α atoms of channel-lining residues (S_i, S_m, and S_o) of each monomer, were constructed at 0.25 Å bin width from the last 8 ns of the simulations. The PMF profiles were calculated from the histograms and were normalized with respect to the PMF at S_i or S_o sites.

8.2.6 Experimental Methods

Our collaborators, Walter F. Boron lab at Case Western University, Cleveland, performed all the physiological measurements on human UT-B. i.e., urea uptake as well as water and ammonia permeabilities. Please see [251] for the details of the experimental methods.

8.3 Results and Discussion

8.3.1 Structural Basis and Energetics of NH_3 and Water Conduction

Physiological experiments have indicated performed by our collaborators at Boron Lab that human UT-B increased the influx of NH_3 and water, but not CO_2. To investigate the involvement of the monomeric pores in the conduction of water and NH_3 molecules as indicated by the physiological experiments performed by Boron Lab and to characterize the pathway, mechanism, and energetics associated with such conduction events, we performed MD simulations on a membrane-embedded model of bUT-B. Here we report free-energy (PMF) profiles associated with permeation of water and NH_3 through the monomeric pores (Fig. 8.1). We previously identified two major barrier regions against

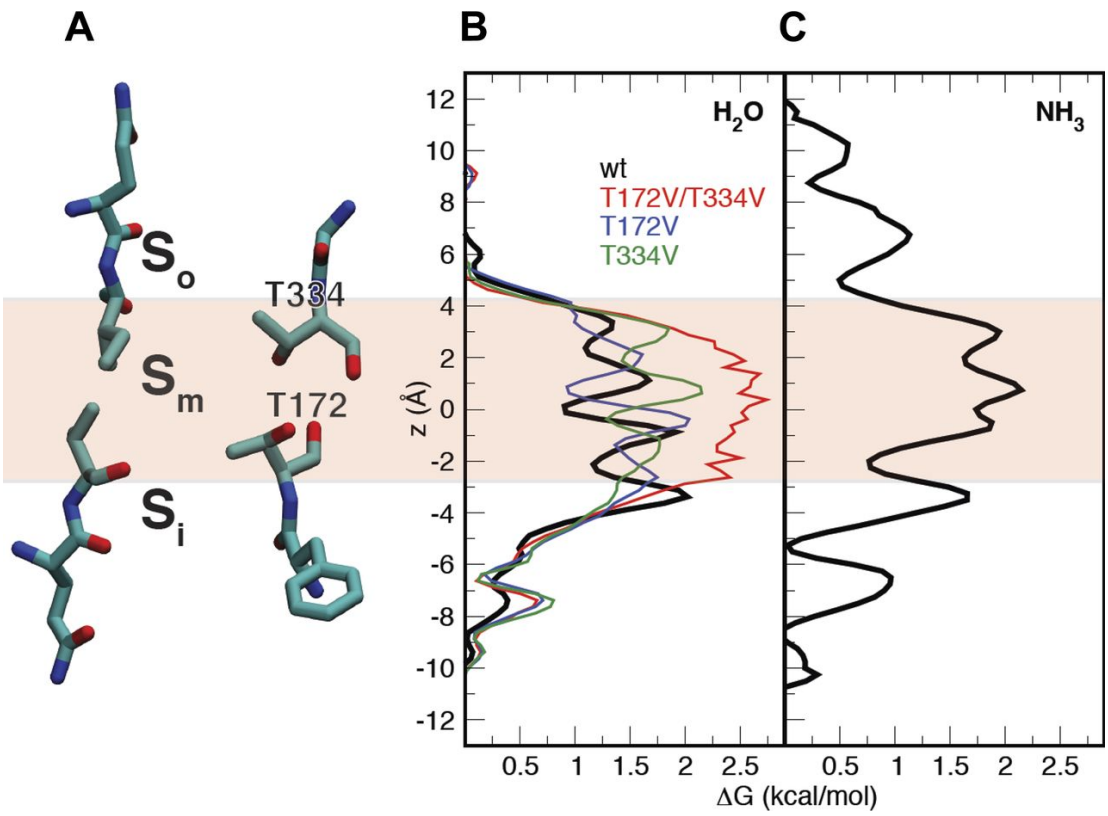

Figure 8.1: Structure of the selectivity filter and potential of mean force (PMF) profiles of NH_3 and water permeation through the monomeric pore of bUT-B. (A) key channel-lining residues around the selectivity filter forming the major substrate binding sites (S_i, S_m, and S_o) are shown in licorice. Residue numbers refer to bUT-B. (B) PMF profile for water calculated from unbiased simulations of wild-type bUT-B, double mutant (T172V/T334V), and single mutants (T172V and T334V). (C) PMF profiles calculated using umbrella sampling simulations for permeation of NH_3 through monomeric pores.

urea permeation around the S_m site (located between T172 and T334 of bUT-B, corresponding to T177 and T339 of hUT-B, respectively; see Fig. 8A), with heights of ∼5 kcal/mol at T172, and ∼4 kcal/mol at T334, which would allow for urea transport through the pores at room temperature (Fig. 7.4).

We used a similar approach here, based on MD simulations and US calculations to characterize the mechanism and energetics of permeation of water and NH_3 through the monomeric pores. We calculated the PMF for water permeation using the distribution of water molecules obtained from unbiased MD simulations and calculated the PMF for NH_3 permeation using US calculations. The free-energy profiles associated with the permeation of these small molecular species through the monomeric pores of bUT-B are shown in Fig. 8.1. In both cases, the S_m region, which constitutes the narrowest portion

of the pore, seems to present the main energy barrier, although the barrier is rather small and easily surmountable at room temperature, allowing permeation of these species through the monomeric pores. Our simulations reveal that the bUT-B monomeric pores are permeable to water, a phenomenon that was suggested to play a role in the mechanism of urea transport [49]. The PMF profile for water permeation exhibits a barrier of only $\sim$2 kcal/mol at the S_m site (Fig. 8.1B). The PMF profile for NH_3 permeation reveals a maximum free-energy barrier of only $\sim$2.2 kcal/mol at the S_m site (Fig. 8.1C), clearly indicating that NH_3 can permeate the channel very efficiently. Three local minima at the S_m site qualitatively match between the PMF profiles of water and NH_3. The simulation predicts that the T172V/T334V double mutant results in a significant decrease in water permeation through the S_m site (Fig. 8.1B) by constricting the pore at that region [49]. On the other hand, the single mutants raise the barrier for water permeation around the position of the mutation, still preserving some of the structure at the S_m site (Fig. 8.1B). Small changes of the pore size by hydrophobic residue mutations might be sufficient also for attenuating the influx of polar urea and NH_3.

Structural analysis of the central (trimeric) cavity, together with our observation that no water penetrated into this region during the simulations, strongly suggest that the central cavity is an unlikely pathway for ammonia conduction. We attribute this observation to the hydrophobicity and physical occlusion of this cavity to both sides of the membrane.

8.4 Summary

This study confirmed that UT-B conducts water and showed that UT-B also conducts NH_3. The classical view had been that all transport of gases across biological membranes occurs as the gas simply dissolves in the lipid phase of the cell membrane and then moves by simple diffusion through the lipid. This idea has been challenged since the discovery of the first gas-impermeable membrane [264] and the observation that AQP1 conducts CO_2 [252, 257], making AQP1 the first identified gas channel. Subsequent experimental

work also showed that AQP1 can conduct NH_3 [259] and NO [263], and simulation studies provided a structural basis and putative protein pathways for conduction of these gas species through aquaporins [39, 265, 266].

The second family of gas channels is the Rhesus proteins, which conduct NH_3 [260, 267, 268]. In human red blood cells, RhAG and the rest of the Rh complex are responsible for about half the CO_2 permeability, with AQP1 being responsible for the other half [269]. Thus both AQP1 and RhAG conduct both CO_2 and NH_3. More interestingly, it was experimentally confirmed that, like ion channels, gas channels (i.e., AQPs and Rh proteins) exhibit selectivity for one gas over another (i.e., NH_3 vs. CO_2) [256].

This study demonstrated that UT-B conducts NH_3, which makes the UTs the third family of gas channels, along with the AQPs and Rhs. MD simulations in the light of the experiments suggest that urea, water, and NH_3 all traverse the membrane utilizing the same pathway, presumably the monomeric urea pores.

Chapter 9

Structure and Dynamics of Aquaporin-Water Interactions *

9.1 Introduction

Aquaporins are ubiquitously found in all kingdoms of life, as primary water channels [104]. Aquaporins have evolved selectivity mechanisms to transport water while excluding hydroxide (OH^-) and hydronium (H_3O^+) ions, and preventing proton transport via a Grotthuss mechanism [105] in which protons hop through hydrogen bonded water molecules. It has been well established that aquaporins have six transmembrane α-helices and form tetramers [106–117] Loops B and E bring two half-helices together, forming a seventh pseudo-transmembrane helix that insert from opposite sides of the membrane, which, in turn, place the conserved dual asparagine-proline-alanine (NPA) signature motif at the center of the monomeric pore (Fig. 9.1). The aromatic/arginine selectivity filter (SF), which is the narrowest part of the monomeric pore at the extracellular side, is responsible for the transport specificity [118, 119].

Most studies performed on the aquaporins suggest a variety of mechanisms for proton exclusion: electrostatic repulsion [121, 123–125], configurational barriers [119], and des-

This chapter appeared as a report in Science [50]. Urszula Kosinska Eriksson, Gerhard Fischer, Rosmarie Friemann, Giray Enkavi, Emad Tajkhorshid, Richard Neutze*, "Subangstrom resolution x-ray structure details aquaporin-water interactions". *Science*, 340(6138):1346–1349, **2013**.

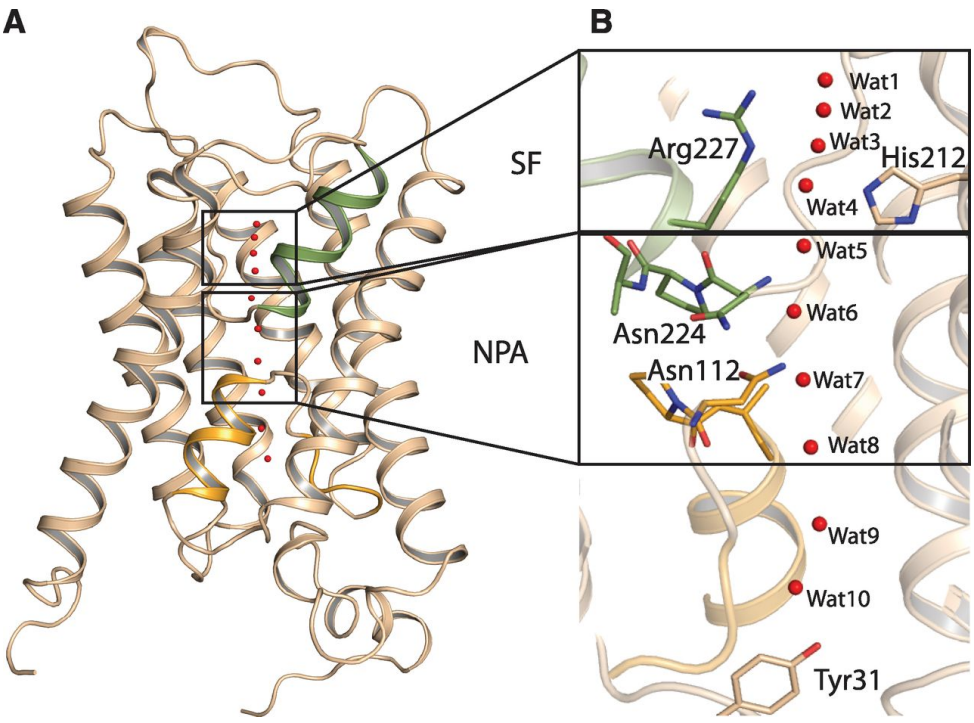

Figure 9.1: Fold of Aqy1. (A) The six transmembrane helices and the seventh pseudo-transmembrane helix formed by loops B (orange) and E (green). (B) Water molecule positions within the channel (red spheres). The dual-NPA-aquaporin signature motif (bottom box) and the SF (top box) are highlighted. [a]

[a]This figure was prepared by the other co-authors of the report published in Science [50].

olvation penalties [126]. The common feature of all these mechanisms is that the NPA region, where the two macrodipoles of two half-helices meet, is the main barrier against proton transport [119–126], by restraining the orientation of the water molecule at this region to disfavor Grotthuss mechanism [119, 120]. However, mutations within NPA motifs that neutralizes its electrostatic barrier result in leaking of Na^+, but not protons [127,128], while mutations within the SF allows protons conduction [129, 130].,

To examine the underlying mechanism of facilitated, selective water transport, and the role of the SF in excluding protons, our collaborators, Richard Neutze lab (University of Gotenburg, Sweden), first, optimized crystals of Aqy1 [109], the sole aquaporin of *Pichia pastoris*, and determined its crystal structure to 0.88 Å resolution. We employed molecular dynamics simulations to elucidate the water dynamics in the channel and explain the features seen in the crystal structure.

9.2 Methods

9.2.1 System Preparation

The tetramer of wild-type Aqy1 was constructed by superimposing the C_α of the monomer onto the C_α of individual subunits of 2w2e structure [109] obtained from the OPM (Orientations of Proteins in Membranes) database [177]. Each monomer was chosen to have a different combination of alternative conformations for the side chains present in the crystal structure, in order to assess the potential effect of these on the conclusions. The protonation states of side chains were determined with MolProbity [173, 174]. After removing all the detergent molecules, CHARMM-GUI [270, 271] was used to embed the tetramer into a POPE bilayer ($\sim$100 by 100 Å^2), generate pore water, and solvate and ionize the system with 100 mM NaCl, resulting in system of $\sim$100 by 100 by 100 Å^3 dimension and 97,791 atoms (wild-type Aqy1). The system was then energy minimized for 5000 steps and simulated for 3 ns under constant 1 atm pressure and 300 K temperature (NPT), while all heavy atoms were harmonically restrained ($\text{k} = 5\,\text{kcal/mol/Å}^2$) to allow relaxation and packing of the lipids against the protein.

9.2.2 Simulation Systems

After relaxation of the lipids and the system was further energy minimized for 5000 steps and simulated for 20 ns.

9.2.3 Simulation Protocols

The simulation was performed using NAMD 2.8b3 [163] with CHARMM27 force field with ϕ/ψ cross term map (CMAP) corrections for the protein [161] and CHARMM36 all-atom additive parameters for lipids [162]. Water was modeled as TIP3P [179]. All simulations were maintained at 1.0 atm using the Nos-Hoover Langevin piston method [180,181] and at 300 K using Langevin dynamics with a damping coefficient of 0.5 ps^-1 applied to all non-hydrogen atoms. Short-range interactions were cut off at 12 Å with a smoothing function

applied after 10 Å, and long-range electrostatic forces were calculated using the particle mesh Ewald (PME) method [182] at a grid density of >1 $Å^{-3}$. Bonded, non-bonded, and PME calculations were performed at 2-, 2-, and 4-fs intervals, respectively.

9.2.4 Specific Analysis

All analyses were performed on the last 15 ns of the production simulation using VMD [164].

Velocity Correlations of Water Molecules inside Water Pore

The velocity correlations as a function of z between the neighboring hydrogen bonded water molecules are calculated using

$$corr(z) = \frac{\langle \delta(z - z_1(t))(\Delta z_1 \cdot \Delta z_2) \rangle}{\sqrt{\langle \delta(z - z_1(t)) \Delta z_2^2 \rangle} \sqrt{\langle \delta(z - z_1(t)) \Delta z_2^2 \rangle}},$$

$$\Delta z_1 = z_1(t + \Delta t) - z_1(t), \qquad \Delta z_2 = z_j(t + \Delta t) - z_2(t),$$

where z is the z-coordinate, z_1 and z_2 are the indices of two neighboring water molecules with $z_2 > z_1$, δ is the Dirac delta function, and $\Delta t = 10$ ps.

Clustering Water Molecules inside the Water Pore

We binned the z-positions of the water oxygen atoms inside the channel in any given frame of the simulation using the z-position of the crystal structure water oxygen atoms as the centers of the bins. That is, each water molecule inside the channel is assigned to one of the 10 crystal structure water molecule for each frame. We, then, counted each configuration (a binary list composed of 10 numbers; 0 for no water presence at the position of the crystal structure and 1 for a water oxygen present) obtained in the simulations and calculated the probabilities by dividing the number of each configuration by the number of frames.

Co-occupancy Analysis of Water Oxygens inside the Water Pore

The co-occupancy analysis was performed by binning the z-coordinates of water oxygen atoms at each frame using a bin size of 0.25 Å. Then, for each bin, we sum all the histograms, in which that particular bin has an occupancy of 1 or more. This results in a matrix that is symmetric around the diagonal that extends from lower right to upper right. We then normalized the histograms by dividing each row by the value at the diagonal in order to get the conditional probability, $p(z|z')$. We, then, converted the probabilities to $\Delta\Delta G$ using the equation $\Delta\Delta G = -k_B T \log p(z|z') - (-k_B T \log(p(z|bulk))$, where $p(z|z')$ is the conditional probability of z being occupied given z' is occupied and $p(z|bulk)$ is the conditional probability of z normalized by average values in bulk water.

9.2.5 Summary of Experimental Methods

Richard Neutze lab (University of Gotenburg, Sweden) overexpressed, purified, and crystallized the only aquaporin found in *Pichia pastoris*, Aqy1 (PDB ID 2w2e). For the details of the experimental methods and conditions please refer to [50].

9.3 Results and Discussion

9.3.1 Water Interactions and Structure near NPA Motif

Figure 9.2 shows two types of electron densities: the $2mF_{\text{obs}} - DF_{\text{calc}}$, which is interpreted as the model electron density, and the $mF_{\text{obs}} - DF_{\text{calc}}$, which reveals the part of the observed density that are not accounted in the model [272]. At sub-angstrom resolution, the conformations of the two NPA asparagine residues (N112 and N224) could be uniquely determined since electron density is delocalized across the carbon-oxygen double bond of these side chains, whereas that for the side-chain nitrogen atom is more localized (Fig. 9.2). Moreover, $mF_{\text{obs}} - DF_{\text{calc}}$ reveals electron densities associated with all four proton-donor interactions of the N_δ atoms of the dual-NPA asparagine residues. Hydrogen bond donor interactions of N224:N_δ to water molecule 6 (Wat6) (Fig. 9.2A) and to the carbonyl

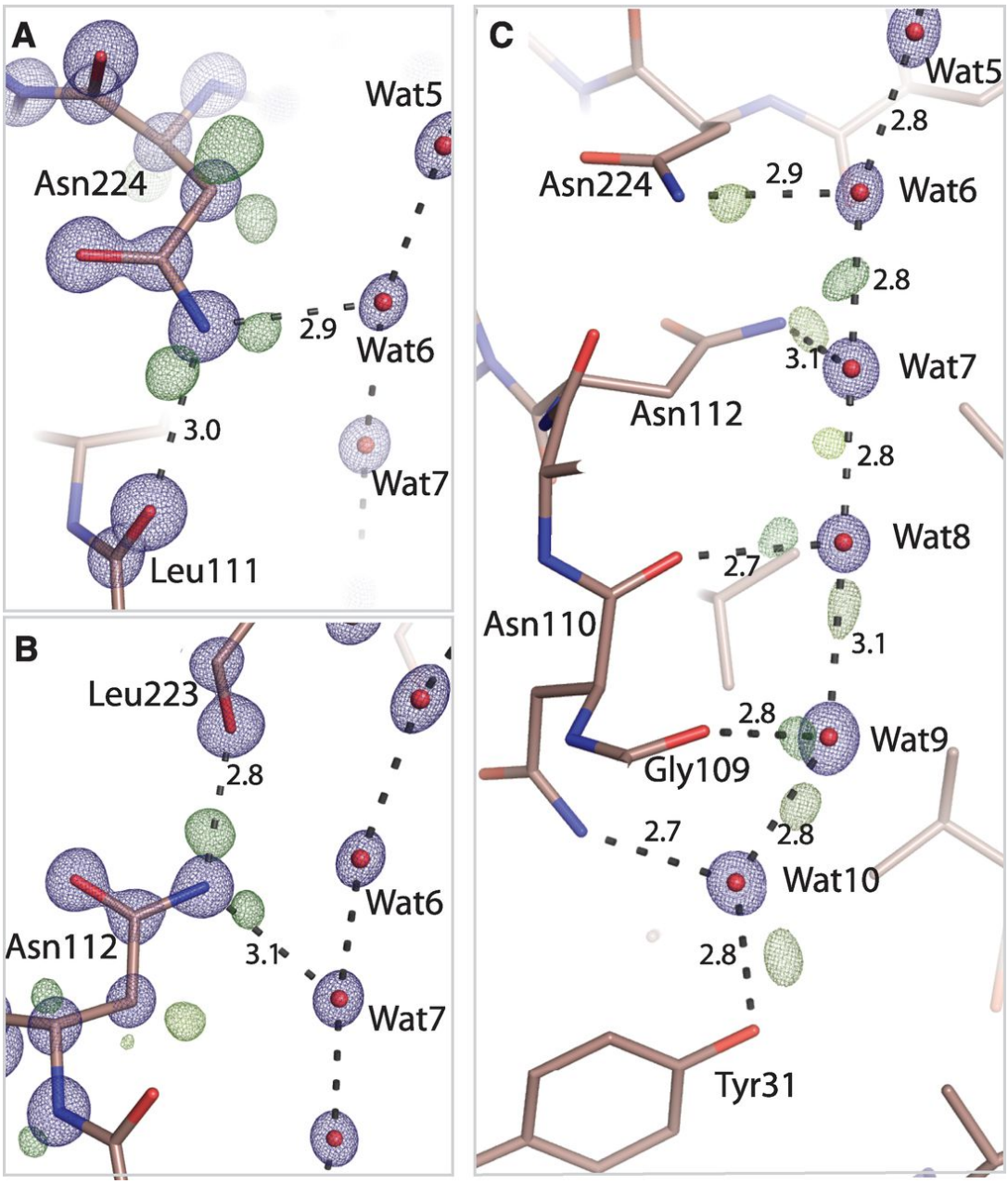

Figure 9.2: Electron density within the NPA and cytoplasmic regions of the Aqy1. The contour levels are colored as follows: blue at 4.3 e/Å^3; green at 0.33 e/Å^3; yellow-green at 0.26 e/Å^3; brown-green at 0.59 e/Å^3, dark green at 0.39 e/Å^3, and light green at 0.15 e/Å^3. (A) $2mF_{\rm obs} - DF_{\rm calc}$ and $mF_{\rm obs} - DF_{\rm calc}$ electron densities associated with (A) N224 and (B) N112. Delocalized $2mF_{\rm obs} - DF_{\rm calc}$ density connects the dual-NPA asparagine C_γ and O_δ atoms, whereas that associated with N_δ is more localized. Residual $mF_{\rm obs} - DF_{\rm calc}$ density indicates hydrogen bond donor interactions with passing water molecules. (C) The $2mF_{\rm obs} - DF_{\rm calc}$ electron density illustrates the position of water molecules, and $mF_{\rm obs} - DF_{\rm calc}$ residual electron density indicates water hydrogen bond interactions within the aquaporin channel. [a]

[a]This figure was prepared by the other co-authors of the report published in Science [50].

oxygen of L111, as well as hydrogen bond donor interactions of N_δ of N112 to Wat7 (Fig. 9.2B) and to the carbonyl oxygen of L223, are all resolved in the crystal structure. These observations confirm that hydrogen bond donor interactions from the NPA motifs constrain the orientation of passing water molecules [115,119,120]. No modeled water has hydrogen bond interactions with both NPA asparagines, as is often depicted [115, 119, 121,124,125]. Therefore, based on the crystal structure, a water molecule at this position cannot be the critical ingredient preventing Grotthuss proton transport.

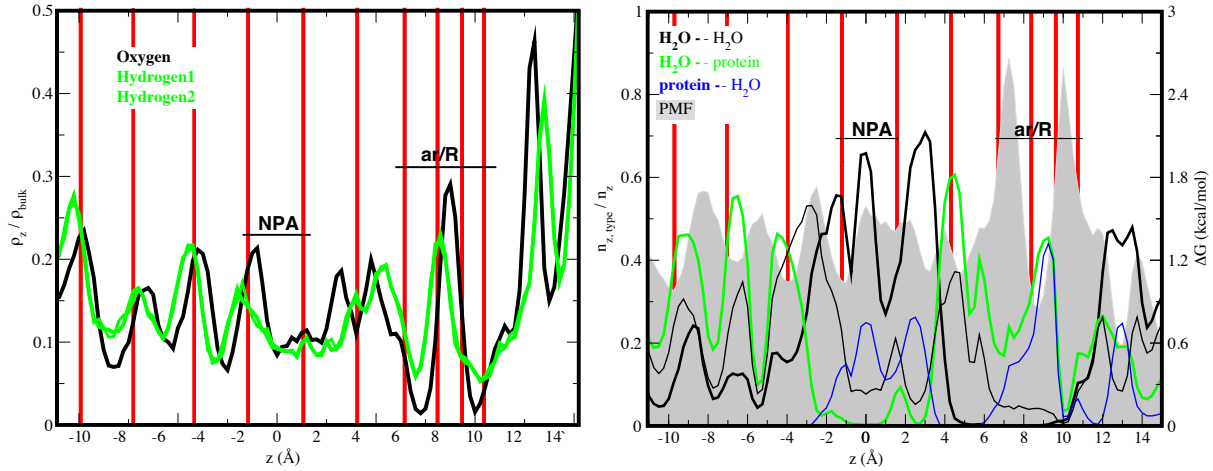

Figure 9.3: Hydrogen bonding patterns of water in MD simulations. Red lines indicate the positions of the crystal water molecules. (left) Densities of the water O atoms (black) and H atoms (green) normalized to their densities in bulk water. (right) Hydrogen bond donor–acceptor densities in the monomeric channel of Aqy1. Number of hydrogen bonds as a function of the channel axis are counted and classified. Thick black line represents water donating hydrogen bonds to water, thin black line represents water accepting hydrogen bonds from water. Thick green line represents water donating hydrogen bonds to protein side chains. Thin blue line represents water accepting hydrogen bonds from protein side chains. All counts ($n_{z,type}$) are divided by the total number of all donors and acceptors in that value of z (n_z). Potential of mean force (PMF) profile for the water O atoms is shown as a gray shade.

Previous MD simulations have predicted that water molecules adopt a bipolar orientation in the two halves of the channel, such that proton-donor interactions systematically point away from the NPA region, which disfavors Grotthuss proton exchange [115, 119, 120]. In the cytoplasmic half-channel, electron densities reveal that Wat8 donates a hydrogen bond to Wat9 (Fig. 9.2C); Wat9 to Wat10; Wat10 to the hydroxyl group of Y31, and weaker residual density suggests a hydrogen bond from Wat7 to Wat8. These observations are confirmed by the our MD simulations. Figure 9.3 shows the densities of the O and H atoms of water in the channel. Dipole moments of the water molecules in the

cytoplasmic half channel are oriented so that they donate hydrogen bonds to each other in the aforementioned order (Wat8-Wat9-Wat10). Moreover, water molecules also donate hydrogen bonds to the protein side chains in this region (see Fig. 9.3, green line on the left panel).

A positive electron density peak located almost exactly between Wat6 and Wat7 suggests that these two waters donate a hydrogen bond to each other with approximately equal probability, which illustrates that hydrogen bond directionality is not imposed at the very center of the channel where the positive electrostatic potential is at its maximum. MD simulations show that the water molecules in the NPA region accept hydrogen bonds from the protein (Fig. 9.3, blue line), and donate hydrogen bonds mostly to either half of the channel. Indeed, the simulations show that Wat6 may not have been captured in a stable position in the crystal structure. Since Wat6 does not correspond to a minimum on the PMF (Fig. 9.3, dark gray shade on the right) nor the volumetric densities (Fig. 9.4B), Wat6 represents a relatively unstable water position. A more stable water position indeed exists slightly on the cytoplasmic side and is separated from Wat5 position by a small barrier. The clustering analysis also shows that >60% of the time, cytoplasmic-half channel (Wat5 to Wat10) is fully occupied, while >20% of the time, the Wat6 position is unoccupied (Fig. 9.5) as reflected in our density plots and its larger anisotropic motions in the crystal structure (Fig. 9.4B). Nevertheless, all assigned interactions are consistent with the previously proposed bipolar distribution of water-water hydrogen bonds in the cytoplasmic and extracellular halves of the channel [115, 119, 120], and this polarization does not depend on the presence of a water molecule that simultaneously accepts hydrogen bonds from both NPA asparagines.

MD simulations reproduce the crystallographic positions of Wat5 and Wat7 to Wat10 as the most probable channel water positions at any given moment (Fig. 9.4B), The hydrogen bond connectivity of water molecules is significantly perturbed at both the NPA region of the channel, preventing Grotthuss proton transport (Fig. 9.7). Within the NPA region, this disruption is accompanied by a reduction in the correlated motion of water molecules (Fig. 9.4A), because one water molecule (corresponding to Wat6) rapidly

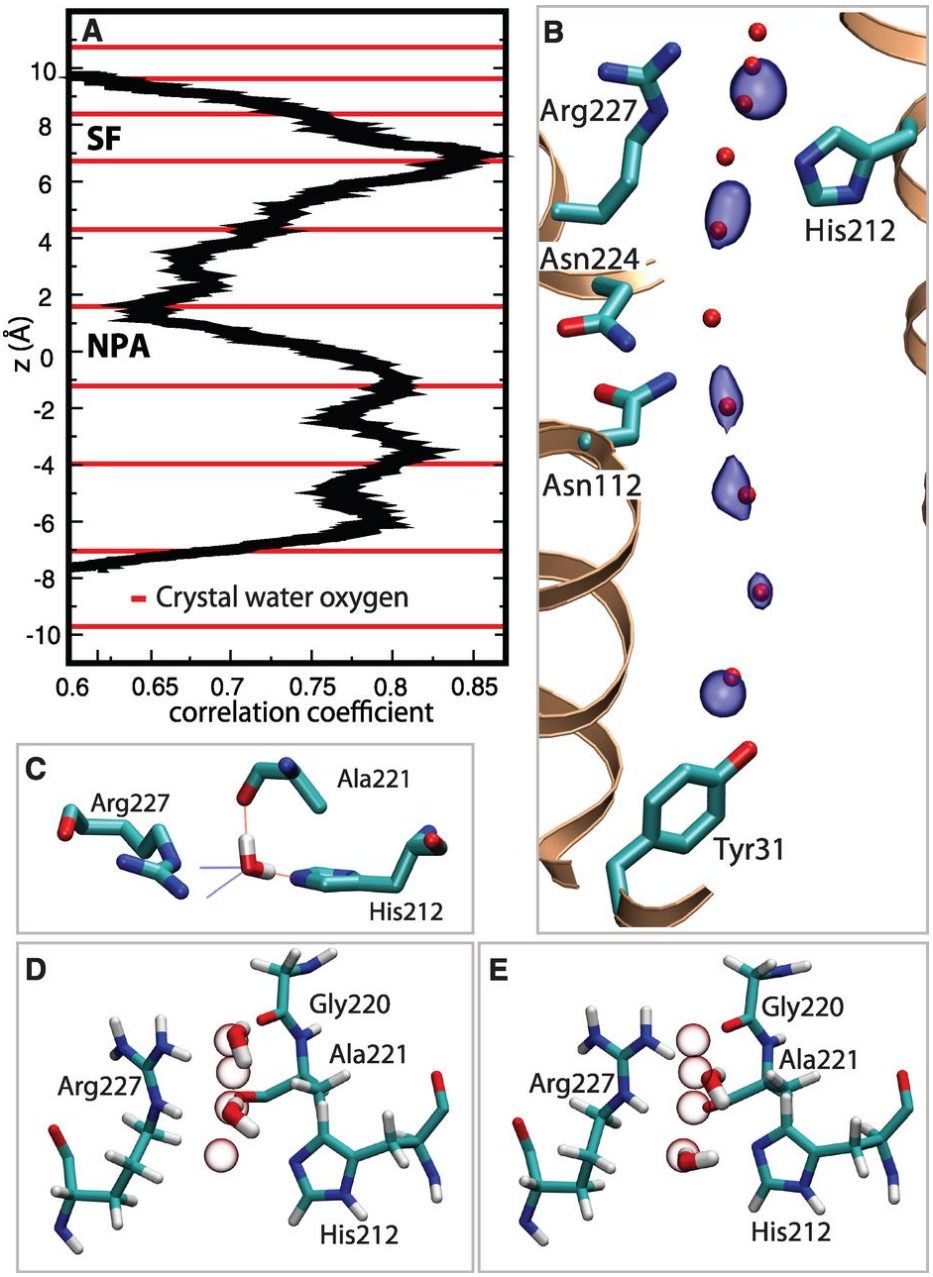

Figure 9.4: (A) Plot illustrating the correlation of movements of adjacent water molecules. Strongly correlated movements arise in the SF and in the cytoplasmic half of the channel. The positions of oxygen atoms of crystallographic waters are indicated as red lines. For the details of the analysis, see Section 9.2.4 (B) Surface representation (blue) of the most probable positions of water molecules (averaged over the final 15-ns of a 20-ns trajectory) superimposed on the crystallographic water positions (red spheres). (C) Snapshot of a water molecule within the SF with all four hydrogen bond interactions occupied. (D and E) Snapshots corresponding to Wat1/Wat3 and Wat2/Wat4 water occupancy of the SF, which indicate the pair-wise movement of water molecules in this region. All four closely spaced SF crystallographic water positions are indicated as white spheres.

alternates its hydrogen bond interactions between neighboring waters on either half of the channel. This finding is consistent with the crystal structure, because residual electron density between Wat6 and Wat7 is significantly weaker than that observed between Wat8, Wat9 and Wat10, and no residual density is visible between Wat5 and Wat6 to suggest a well-defined hydrogen bond.

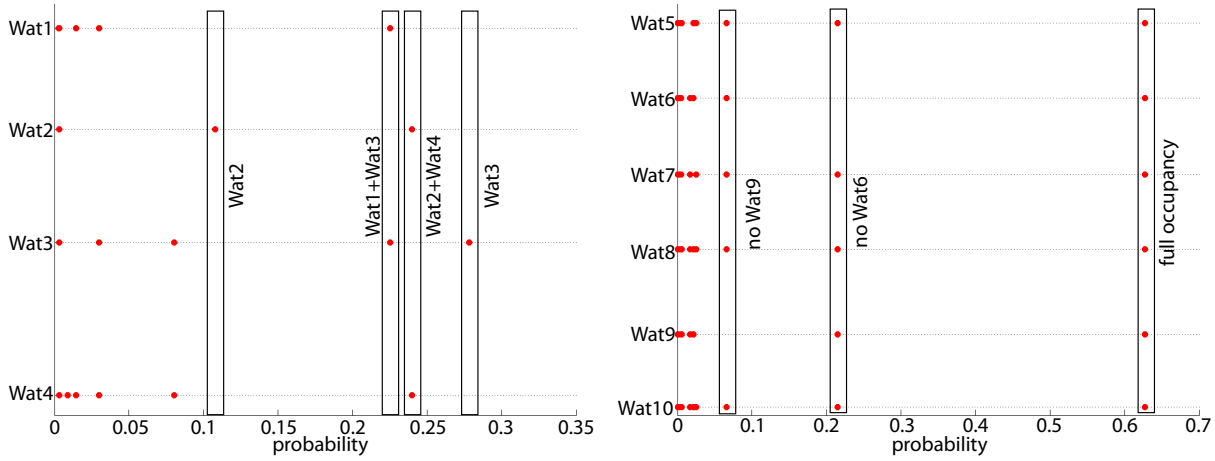

Figure 9.5: The probabilities of configurations of water oxygens in the channel. For the details of the analysis, see Section 9.2.4 The red dots represent oxygen atoms. The horizontal lines indicate the position of the water oxygens in the crystal structure. The horizontally separated bands of red dots represent each configuration sampled in the simulations. For each frame, the positions of water molecules inside the channel is matched to the closest crystal structure water oxygen atom for clustering purposes. The x-axis shows the probability of a particular water configuration sampled in the simulations. Water oxygens in the SF are shown on the left, and those below the SF are shown on the right.

9.3.2 Water Interactions and Structure in Selectivity Filter

Placement of two additional water molecules and unrestrained refinement of their crystallographic occupancies and B-factors revealed four stable water positions within the SF with complementary occupancies of 66% (Wat2/Wat4) (Fig. 9.2) and 34% (Wat1/Wat3). Given their small separation (∼1.5 Å) (Fig. 9.2), it is not possible for all four water positions to be occupied simultaneously. MD simulations predict only one high-probability water position within the SF (between Wat2 and Wat3) when averaged over an entire trajectory. However, seeing that specific snapshots capture the crystallographic water configurations (Fig. 9.4D–E), we extended our analysis of the trajectories to extract more information about the water positions.

The clustering analysis shows that the SF region is mostly doubly occupied with Wat2/Wat4 and Wat1/Wat3 configurations having a total of >45% probability. However, only Wat3 occupying the SF appears as the most probable configuration in our simulations (Fig. 9.3 and Fig. 9.4B). This is most probably due to Wat1 position being close to the opening of the channel and less stable. Therefore, only Wat3 configuration can indeed be representation of Wat1/Wat3 configuration.

Both high- and low-occupancy SF water positions have hydrogen bond interactions with the same atoms: with N_η of R227 and the carbonyl oxygen of G220 for Wat1 and Wat2; and N_δ of R227, N_ϵ of H212, and the carbonyl oxygen of A221 for Wat3 and Wat4 (Fig. 9.6). MD snapshots illustrate how this geometry achieves exceptional water selectivity, because all four hydrogen bond donor and acceptor interactions are filled as water moves through the SF (Fig. 9.4C). The presence of four closely spaced water-selective sites optimizes the aquaporin SF's ability to discriminate water from other small molecules. Hydroxide ions, in particular, suffer a geometric penalty, because they cannot simultaneously donate hydrogen bonds to the backbone hydroxyl of A221 and to N_ϵ of H212. Conversely, all hydrogen bond interactions are distorted from ideal water geometry, and this avoids binding water too tightly, such that efficient transport is compromised.

In contrast to the NPA region where velocity correlations of neighboring waters are reduced, water molecules move pair-wise through the SF in a highly correlated manner (Fig. 9.4A), and their connectivity to water molecules outside of the SF is weak. Correlated motions within the cytoplasmic half of the channel emerge from well-defined water-water hydrogen bond interactions (Fig. 9.4C), whereas the correlations observed within the SF appear to be dictated by protein-water hydrogen bond interactions. In addition to electrostatic effects, a disruption of the highly constrained SF water structure (Fig. 9.6) can explain why the mechanism of proton exclusion is sensitive to mutation of the conserved SF arginine and histidine residues [129, 130]. These findings illustrate how evolution has fine-tuned the water channel geometry to optimize protein function, suppressing proton transport without compromising water flux.

We also calculated the co-occupancies of z-positions along the channel. The co-

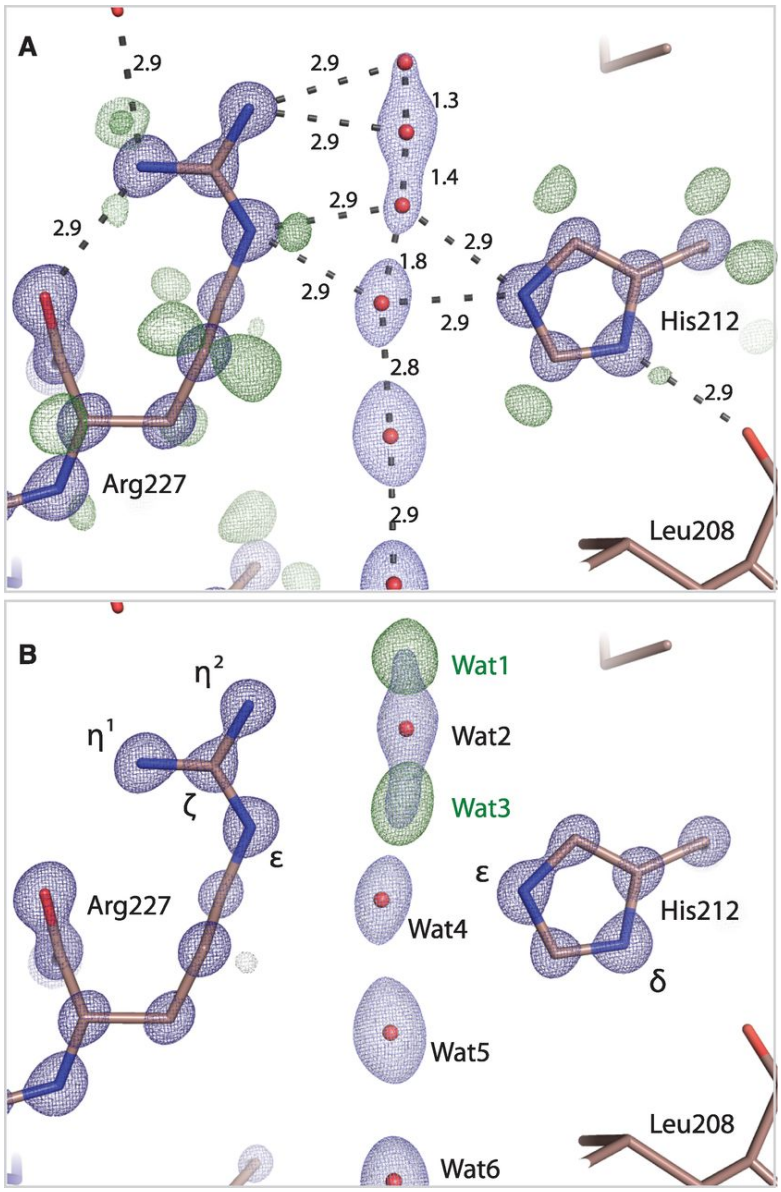

Figure 9.6: Electron density within the Aqy1 SF. (A) The $2mF_{\text{obs}} - DF_{\text{calc}}$ (dark blue contoured at $4.3\,\text{e}/\text{Å}^3$ and light blue at $1.9\,\text{e}/\text{Å}^3$) and residual $mF_{\text{obs}} - DF_{\text{calc}}$ (dark green contoured at $0.42\,\text{e}/\text{Å}^3$ and light green contoured at $0.33\,\text{e}/\text{Å}^3$) electron density associated with H212, R227, and water molecules within the SF. Atomic separations (Å) are indicated. Residual $mF_{\text{obs}} - DF_{\text{calc}}$ electron density reveals that N_δ of H212 is protonated, whereas N_ϵ is not. Connected $2mF_{\text{obs}} - DF_{\text{calc}}$ electron density suggests that the R227 covalent bond from C_ζ to N_η is preferentially conjugated. Four closely spaced water molecules are modeled within the SF with complementary occupancy (66% occupancy, positions 2 and 4; 34% occupancy, positions 1 and 3). (B) The $mF_{\text{obs}} - DF_{\text{calc}}$ omit electron density map calculated when waters 1 and 3 are removed from the structural model (dark green contoured at $0.65\,\text{e}/\text{Å}^3$). Positive electron-density features associated with these waters are the strongest within the channel. [a]

[a]This figure was prepared by the other co-authors of the report published in Science [50].

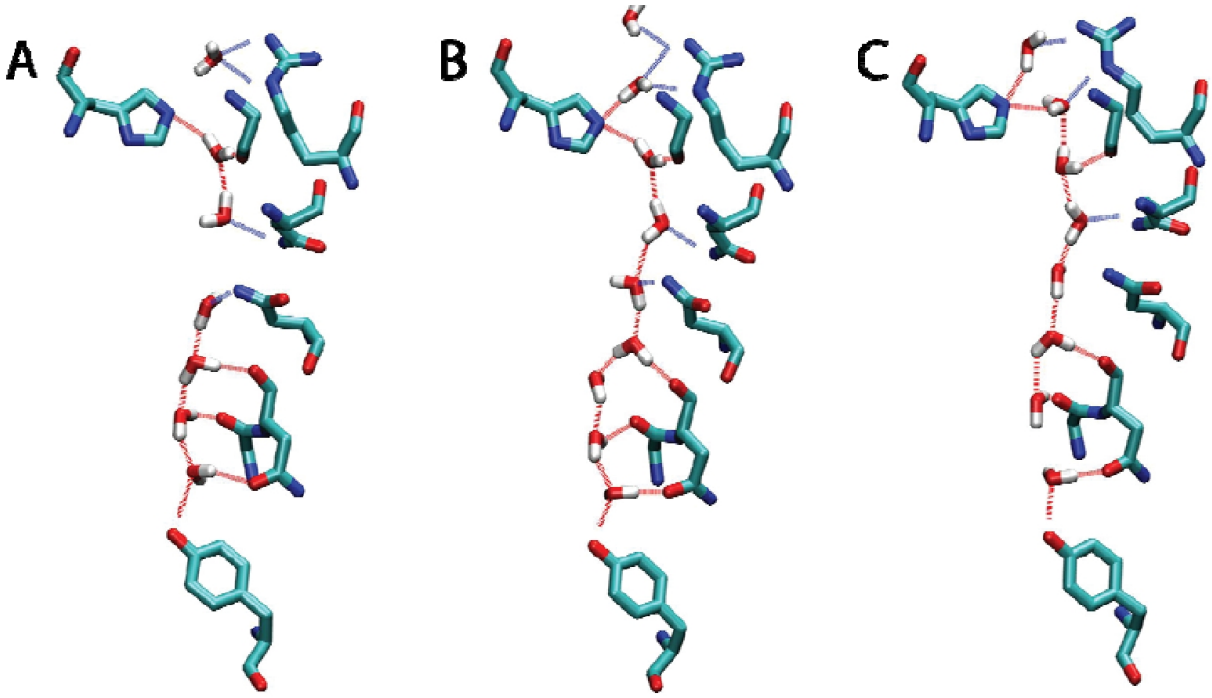

Figure 9.7: Molecular dynamics snapshots illustrating transient water conformations that stretch through the Aqy1 pore. (A–C) Since N_ϵ of H212 is not protonated, the H-bond donor interactions of passing water molecules mean that the geometry is seldom ideal for the exchange of protons within the SF.

occupancy map (Fig. 9.8) shows that the correlations between the co-occupancies are broken at the level of Wat5, confirming the role of NPA in preventing water transport through Grotthuss mechanism. Besides, the map reveals the binary occupancies (Wat1-Wat3/Wat2-Wat4) at the SF. Another information that can be inferred from the map is the correlated motions of the water molecules. While a minimum that extends parallel to the diagonal indicates a correlated motion, the relatively vertical and horizontal minima seen around SF region in relation to Wat5 indicates that the SF water molecules have correlated motion within each other but are uncorrelated with Wat5. When Wat5 is present, it appears that Wat1-Wat2 and Wat3-Wat4 are degenerate. When Wat5, on the other hand, is perturbed towards the SF region, the rest of the water molecules in SF become destabilized. This picture is analogous to the structural mechanism of ion transport through potassium channels, for which partial crystallographic occupancy of four K^+-binding sites within its selectivity filter led to the proposal that potassium ions progress pair-wise through a sequence of four binding sites [273, 274]. As argued for potassium

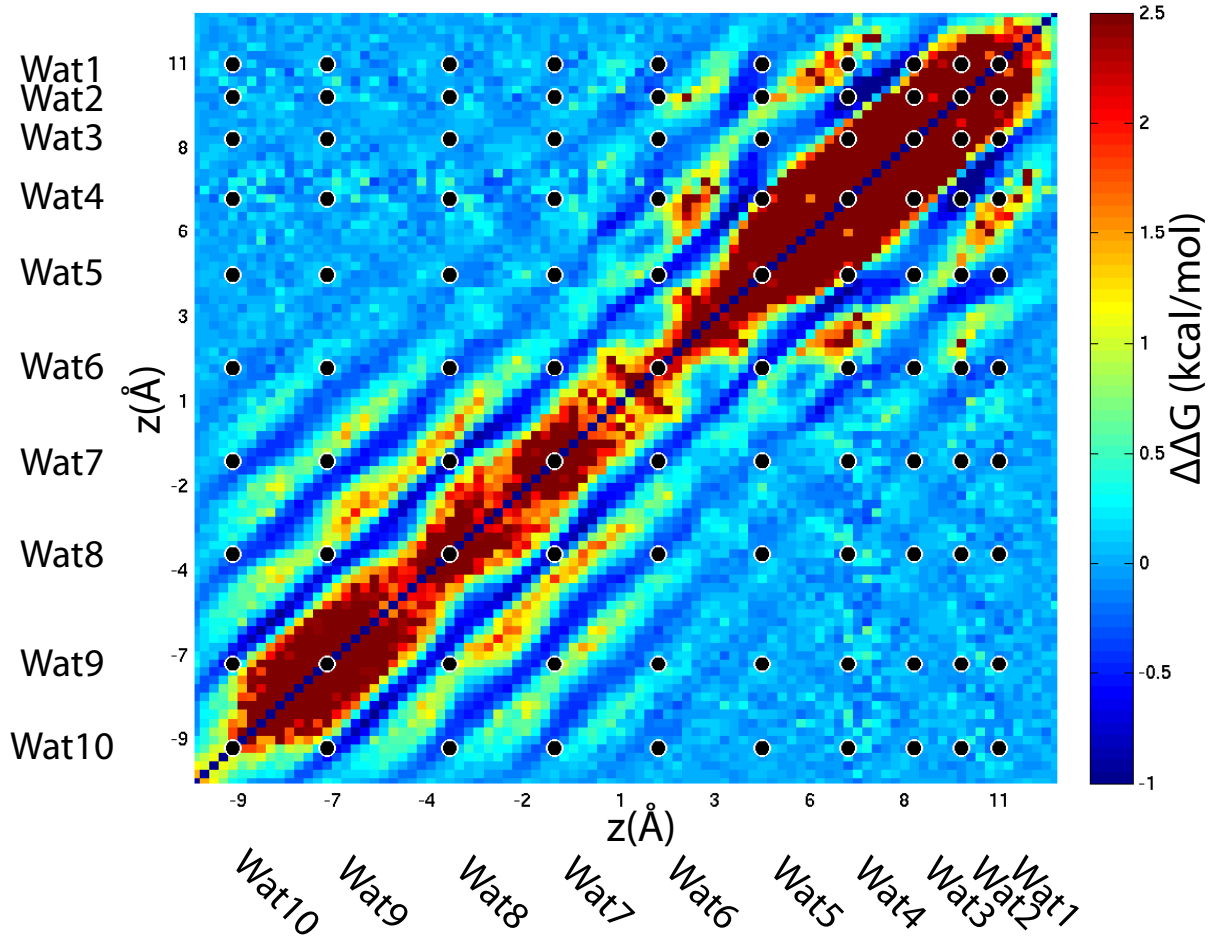

Figure 9.8: The co-occupancy map of the water oxygen atoms in the channel projected onto the z-axis. The black dots represent the position of the oxygen atoms of the crystal water molecules. The cooccupancy map is calculated as described in Section 9.2.4. The map shows whether a water molecule present at z_1 (shown at the diagonal (z_1, z_1)) enhances (dark blue) or diminishes water binding to z_2 (shown at (z_1, z_2) and (z_2, z_1)). That is, for any point on the diagonal, z_1, following along the vertical and horizontal lines points z_2 in red are unlikely to be occupied whereas those in dark blue are more likely to be occupied.

channels [274], the similar crystallographic occupancy of the four aquaporin SF water positions implies very little energy difference between binding configurations Wat1/Wat3 and Wat2/Wat4, which is optimal for maximizing the water conduction rate.

9.4 Summary

The crystal structure of a yeast aquaporin, Aqy1, at 0.88 angstrom resolution is the highest resolution achieved for a membrane protein to date. The crystal structure revealed

the hydrogen bond donor interactions of the NPA motif's asparagine residues to passing water molecules and a polarized water-water H-bond configuration within the channel, the tautomeric states of the SF histidine and arginine residues; and four SF water positions too closely spaced to be simultaneously occupied. MD simulations have provided insight into the dynamics of water movements inside the channel that may account for the filtering of proton permeation. Strongly correlated movements inside the channel break the connectivity of SF waters to other water molecules within the channel and prevent proton transport via a Grotthuss mechanism.

Chapter 10

Bibliography

[1] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, Worth Publishers, 3rd edition, **2000**.

[2] S. A. Shaikh *et al.*, "Capturing functional motions of membrane channels and transporters with molecular dynamics simulation". *Journal of Computational and Theoretical Nanosciences*, 7:2481–2500, **2010**.

[3] J. A. McCammon, B. R. Gelin and M. Karplus, "Dynamics of folded proteins". *Nature*, 267:585–590, **1977**.

[4] L. Celik, B. Schiott and E. Tajkhorshid, "Substrate binding and formation of an occluded state in the leucine transporter". *Biophysical Journal*, 94:1600–1612, **2008**.

[5] C. J. Law *et al.*, "Structural basis of substrate selectivity in the glycerol-3-phosphate:phosphate antiporter GlpT". *Biophysical Journal*, 97:1346–1353, **2009**.

[6] J. Li and E. Tajkhorshid, "Ion-releasing state of a secondary membrane transporter". *Biophysical Journal*, 97:L29–L31, **2009**.

[7] S. A. Shaikh and E. Tajkhorshid, "Potential cation and H^+ binding sites in acid sensing ion channel-1". *Biophysical Journal*, 95:5153–5164, **2008**.

[8] Z. Huang and E. Tajkhorshid, "Dynamics of the extracellular gate and ion-substrate coupling in the glutamate transporter". *Biophysical Journal*, 95:2292–2300, **2008**.

[9] P.-C. Wen and E. Tajkhorshid, "Dimer opening of the nucleotide binding domains of ABC transporters after ATP hydrolysis". *Biophysical Journal*, 95:5100–5110, **2008**.

[10] J. Gumbart, M. C. Wiener and E. Tajkhorshid, "Coupling of calcium and substrate binding through loop alignment in the outer membrane transporter BtuB". *Journal of Molecular Biology*, 393:1129–1142, **2009**.

[11] Z. Huang and E. Tajkhorshid, "Identification of the third Na^+ site and the sequence of extracellular binding events in the glutamate transporter". *Biophysical Journal*, 99:1416–1425, **2010**.

[12] J. Li and E. Tajkhorshid, "A gate-free pathway for substrate release from the inward-facing state of the Na^+-galactose transporter". *Biochimica et Biophysica Acta – Biomembranes*, 1818:263–271, **2012**.

[13] O. Jardetzky, "Simple allosteric model for membrane pumps". *Nature*, 211(5052):969–970, **1966**.

[14] H. Krishnamurthy, C. L. Piscitelli and E. Gouaux, "Unlocking the molecular secrets of sodium-coupled transporters". *Nature*, 459:347–355, **2009**.

[15] K. P. Locher, "Structure and mechanism of ATP-binding cassette transporters". *Philosophical Transactions of the Royal Society of London B (Biological Sciences)*, 364(1514):239–245, **2009**.

[16] O. Boudker and G. Verdon, "Structural perspectives on secondary active transporters". *Trends in Pharmacological Sciences*, 31:418–426, **2010**.

[17] L. R. Forrest, R. Krämer and C. Ziegler, "The structural basis of secondary active transport mechanisms." *Biochimica et Biophysica Acta – Bioenergetics*, 1807(2):167–188, **2011**.

[18] I. C. West, "Ligand conduction and the gated-pore mechanism of transmembrane transport". *Biochimica et Biophysica Acta*, 1331:213–234, **1997**.

[19] L. J. DeFelice, "Transporter structure and mechanism". *Trends in Neurosciences*, 27:352–359, **2004**.

[20] A. Yamashita *et al.*, "Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters". *Nature*, 437:215–233, **2005**.

[21] H. Krishnamurthy and E. Gouaux, "X-ray structures of LeuT in substrate-free outward-open and apo inward-open states." *Nature*, 481:469–474, **2012**.

[22] S. Weyand *et al.*, "Structure and molecular mechanism of a Nucleobase-Cation-Symport-1 family transporter". *Science*, 322:709–713, **2008**.

[23] T. Shimamura *et al.*, "Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1". *Science*, 328(5977):470–473, **2010**.

[24] M. L. Oldham *et al.*, "Crystal structure of a catalytic intermediate of the maltose transporter". *Nature*, 450(7169):515–521, **2007**.

[25] D. Khare *et al.*, "Alternating access in maltose transporter mediated by rigid-body rotations". *Molecular Cell*, 33(4):528–536, **2009**.

[26] M. L. Oldham and J. Chen, "Crystal structure of the maltose transporter in a pretranslocation intermediate state". *Science*, 332(6034):1202–1205, **2011**.

[27] M. L. Oldham and J. Chen, "Snapshots of the maltose transporter during ATP hydrolysis". *Proceedings of the National Academy of Sciences, USA*, 108(37):15152–15156, **2011**.

[28] D. Yernool *et al.*, "Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*". *Nature*, 431:811–818, **2004**.

[29] O. Boudker *et al.*, "Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter". *Nature*, 445:387–393, **2007**.

[30] N. Reyes, C. Ginter and O. Boudker, "Transport mechanism of a bacterial homologue of glutamate transporters". *Nature*, 462:880–885, **2009**.

[31] G. Verdon and O. Boudker, "Crystal structure of an asymmetric trimer of a bacterial glutamate transporter homolog". *Nature Structural & Molecular Biology*, 19:355–357, **2012**.

[32] F. Khalili-Araghi *et al.*, "Molecular dynamics simulations of membrane channels and transporters". *Current Opinion in Structural Biology*, 19:128–137, **2009**.

[33] P. Stansfeld and M. Sansom, "Molecular simulation approaches to membrane proteins". *Structure*, 19(11):1562–1572, **2011**.

[34] E. Perozo *et al.*, "Open channel structure of MscL and the gating mechanism of mechanosensitive channels". *Nature*, 418:942–948, **2002**.

[35] G. V. Miloshevsky and P. C. Jordan, "Permeation in ion channels: the interplay of structure and theory". *Trends in Neurosciences*, 27:308–314, **2004**.

[36] S. Choe, "Potassium channel structures". *Nature Reviews*, 3:115–121, **2002**.

[37] D. P. Tieleman *et al.*, "Simulation approaches to ion channel structure-function relationships". *Q Rev Biophys*, 34:473–561, **2001**.

[38] S. Noskov and B. Roux, "Importance of hydration and dynamic on the selectivity of the KcsA and NaK channels". *Journal of General Physiology*, 129:135–143, **2007**.

[39] Y. Wang *et al.*, "Exploring gas permeability of cellular membranes and membrane channels with molecular dynamics". *Journal of Structural Biology*, 157:534–544, **2007**.

[40] J. Hub and B. de Groot, "Mechanism of selectivity in aquaporins and aquaglyceroporins". *Proceedings of the National Academy of Sciences, USA*, 105:1198–203, **2008**.

[41] P. W. Fowler, K. Tai and M. S. Sansom, "The selectivity of K^+ ion channels: Testing the hypotheses". *Biophysical Journal*, 95:5062–5072, **2008**.

[42] S. Kellenberger and L. Schild, "Epithelial sodium channel/degenerin family of ion channels: A variety of functions for a shared structure". *Physiol Rev*, 82:735–767, **2002**.

[43] C. Domene and M. Sansom, "Potassium channel, ions, and water: simulation studies based on the high resolution X-ray structure of KcsA". *Biophysical Journal*, 85:2787–2800, **2003**.

[44] D. P. Tieleman, "Computer simulations of transport through membranes: passive diffusion, pores, channels and transporters". *Clin Exp Pharmacol Physiol*, 33:893–903, **2006**.

[45] W. L. Ash *et al.*, "Computer simulations of membrane proteins". *Biochimica et Biophysica Acta – Biomembranes*, 1666:158–189, **2004**.

[46] P. Grayson, E. Tajkhorshid and K. Schulten, "Mechanisms of selectivity in channels and enzymes studied with interactive molecular dynamics". *Biophysical Journal*, 85:36–48, **2003**.

[47] J. Cohen and K. Schulten, "Mechanism of anionic conduction across ClC". *Biophysical Journal*, 86:836–845, **2004**.

[48] Y. Huang *et al.*, "Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*". *Science*, 301(5633):616–620, **2003**.

[49] E. J. Levin *et al.*, "Structure and permeation mechanism of a mammalian urea transporter". *Proceedings of the National Academy of Sciences*, 109(28):11194–11199, **2012**.

[50] U. Kosinska Eriksson *et al.*, "Subangstrom resolution x-ray structure details aquaporin-water interactions". *Science*, 340(6138):1346–1349, **2013**.

[51] M. H. Saier, Jr. *et al.*, "The major facilitator superfamily." *Journal of Molecular Microbiology and Biotechnology*, 2:257–279, **1999**.

[52] C. J. Law, P. C. Maloney and D.-N. Wang, "Ins and outs of major facilitator superfamily antiporters". *Annual Review of Microbiology*, 62(1):289–305, **2008**.

[53] V. S. Reddy *et al.*, "The major facilitator superfamily (MFS) revisited: The major facilitator family revisited". *FEBS Journal*, 279(11):2022–2035, **2012**.

[54] S. B. Levy, "Active efflux, a common mechanism for biocide and antibiotic resistance". *J Appl Microbiol*, 92:65S–71S, **2002**.

[55] M. J. Lemieux, Y. Huang and D.-N. Wang, "Glycerol-3-phosphate transporter of *Escherichia coli*: Structure, function and regulation". *Research in Microbiology*, 155(8):623–629, **2004**.

[56] L. Bartoloni *et al.*, "Cloning and characterization of a putative human glycerol 3-phosphate permease gene (SLC37A1 or G3PP) on 21q22.3: mutation analysis in two candidate phenotypes, DFNB10 and a glycerol kinase deficiency". *Genomics*, 70(2):190–200, **2000**.

[57] M. J. Lemieux, Y. Huang and D.-N. Wang, "Crystal structure and mechanism of GlpT, the glycerol-3-phosphate transporter from *E. coli*". *Journal of Electron Microscopy*, 54:i43–i46, **2005**.

[58] M. Saidijam *et al.*, "Microbial drug efflux proteins of the major facilitator superfamily". *Current Drug Targets*, 7:793–811, **2006**.

[59] C. F. Higgins, "Multiple molecular mechanisms for multi-drug resistance transporters". *Nature*, 446(7137):749–757, **2007**.

[60] C. J. Law *et al.*, "Kinetic evidence is consistent with the rocker-switch mechanism of membrane transport by GlpT". *Biochemistry*, 46:12190–12197, **2007**.

[61] S. Dang *et al.*, "Structure of a fucose transporter in an outward-open conformation." *Nature*, 467(7316):734–738, **2010**.

[62] L. Sun *et al.*, "Crystal structure of a bacterial homologue of glucose transporters GLUT 1-4". *Nature*, 490:361–366, **2012**.

[63] E. M. Quistgaard *et al.*, "Structural basis for substrate transport in GLUT-homology family of monosaccharide transporters". *Nature Structural & Molecular Biology*, 20:766–768, **2013**.

[64] S. Radestock and L. R. Forrest, "The Alternating-Access mechanism of MFS transporters arises from Inverted-Topology repeats". *Journal of Molecular Biology*, 407(5):698–715, **2011**.

[65] T. J. Crisman *et al.*, "Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats". *Proceedings of the National Academy of Sciences, USA*, 106:20752–20757, **2009**.

[66] K. Khafizov *et al.*, "A study of the evolution of inverted-topology repeats from LeuT-Fold transporters using AlignMe". *Biochemistry*, 49(50):10702–10713, **2010**.

[67] P. J. Henderson *et al.*, "Homologous sugar-transport proteins in microbes and man." *Biochemical Society Transactions*, 21:1002–1006, **1993**.

[68] I. T. Paulsen *et al.*, "Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes." *Journal of Molecular Biology*, 301:75–100, **2000**.

[69] M. H. Saier, Jr., "Families of transmembrane sugar transport proteins." *Molecular Microbiology*, 35:699–710, **2000**.

[70] M. H. S. Jr. *et al.*, "Phylogenetic characterization of novel transport families revealed by genome analyses". *Biochimica et Biophysica Acta*, 1422:1–56, **1999**.

[71] M. H. Saier, Jr., "Families of transmembrane transporters selective for amino acids and their derivatives". *Microbiology*, 146:1775–1795, **2000**.

[72] M.-C. Fann, A. Busch and P. C. Maloney, "Functional characterization of cysteine residues in GlpT, the glycerol 3-phosphate transporter of *Escherichia coli*". *Journal of Bacteriology*, 185:3863–3870, **2003**.

[73] J. Abramson, H. R. Kaback and S. Iwata, "Structural comparison of lactose permease and the glycerol-3-phosphate antiporter: members of the major facilitator superfamily". *Current Opinion in Structural Biology*, 14:413–419, **2004**.

[74] T. Skarzynski *et al.*, "Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin". *Structure*, 4:1465–1474, **1996**.

[75] A. I. Nilsson *et al.*, "Biological costs and mechanisms of fosfomycin resistance in *Escherichia coli*". *Antimicrobial Agents and Chemotherapy*, 47(9):2850–2858, **2003**.

[76] T. J. Silhavy, I. Hartig-Beecken and W. Boos, "Periplasmic protein related to the sn-glycerol-3-phosphate transport system of *Escherichia coli*." *Journal of Bacteriology*, 126(2):951–958, **1976**.

[77] T. Horii *et al.*, "Emergence of fosfomycin-resistant isolates of Shiga-like toxin-producing *Escherichia coli* O26". *Antimicrob Agents Chemother*, 43:789–793, **1999**.

[78] M. J. Lemieux, "Eukaryotic major facilitator superfamily transporter modeling based on the prokaryotic GlpT crystal structure (review)". *Molecular Membrane Biology*, 24(5):333–341, **2007**.

[79] J. Almqvist *et al.*, "Homology modeling of the human microsomal glucose 6-phosphate transporter explains the mutations that cause the glycogen storage disease type ib ". *Biochemistry*, 43(29):9289–9297, **2004**.

[80] L.-Y. Chen *et al.*, "Structure-function analysis of the glucose-6-phosphate transporter deficient in glycogen storage disease type ib". *Human molecular genetics*, 11(25):3199–3207, **2002**, PMID: 12444104.

[81] H. Hiraiwa, "Inactivation of the glucose 6-phosphate transporter causes glycogen storage disease type 1b". *Journal of Biological Chemistry*, 274(9):5532–5536, **1999**.

[82] M. A. Knepper and J. A. Mindell, "Structural biology: Molecular coin slots for urea". *Nature*, 462(7274):733–734, **2009**.

[83] E. J. Levin, M. Quick and M. Zhou, "Crystal structure of a bacterial homologue of the kidney urea transporter". *Nature*, 462(7274):757–761, **2009**.

[84] M. A. Knepper and R. A. Star, "The vasopressin-regulated urea transporter in renal inner medullary collecting duct". *American Journal of Physiology – Renal Physiology*, 259(3):F393–F401, **1990**.

[85] J. M. Sands, "Mammalian urea transporters". *Annual Review of Physiology*, 65(1):543–566, **2003**.

[86] M. A. Hediger *et al.*, "Structure, regulation and physiological roles of urea transporters." *Kidney International*, 49(6):1615–1623, **1996**.

[87] J. D. Klein, M. A. Blount and J. M. Sands, "Urea transport in the kidney", in R. Terjung (Editor), "Comprehensive Physiology", John Wiley & Sons, Inc., Hoboken, NJ, USA, **2011**.

[88] L. Bankir *et al.*, "Direct and indirect cost of urea excretion". *Kidney international*, 49(6):1598–1607, **1996**.

[89] L. Bankir and B. Yang, "New insights into urea and glucose handling by the kidney, and the urine concentrating mechanism". *Kidney International*, 81(12):1179–1198, **2012**.

[90] R. A. Fenton, "Renal phenotype of UT-A urea transporter knockout mice". *Journal of the American Society of Nephrology*, 16(6):1583–1592, **2005**.

[91] B. Yang, "Urea-selective concentrating defect in transgenic mice lacking urea transporter UT-B". *Journal of Biological Chemistry*, 277(12):10633–10637, **2002**.

[92] B. Yang, "Analysis of double knockout mice lacking aquaporin-1 and urea transporter UT-B. evidence for UT-B-facilitated water transport in erythrocytes". *Journal of Biological Chemistry*, 277(39):36782–36786, **2002**.

[93] R. A. Fenton, "Urinary concentrating defect in mice with selective deletion of phloretin-sensitive urea transporters in the renal collecting duct". *Proceedings of the National Academy of Sciences*, 101(19):7469–7474, **2004**.

[94] T. Lei *et al.*, "Role of thin descending limb urea transport in renal urea handling and the urine concentrating mechanism". *AJP: Renal Physiology*, 301(6):F1251–F1259, **2011**.

[95] S. Uchida *et al.*, "Impaired urea accumulation in the inner medulla of mice lacking the urea transporter UT-A2". *Molecular and Cellular Biology*, 25(16):7357–7363, **2005**.

[96] K. Ranade *et al.*, "Genetic variation in the human urea transporter-2 is associated with variation in blood pressure". *Human Molecular Genetics*, 10(19):2157–2164, **2001**.

[97] M. Garcia-Closas *et al.*, "A genome-wide association study of bladder cancer identifies a new susceptibility locus within SLC14A1, a urea transporter gene on chromosome 18q12.3". *Human Molecular Genetics*, 20(21):4282–4289, **2011**.

[98] T. Rafnar *et al.*, "European genome-wide association study identifies SLC14A1 as a new urinary bladder cancer susceptibility gene". *Human Molecular Genetics*, 20(21):4268–4281, **2011**.

[99] M. H. Levin, R. de la Fuente and A. S. Verkman, "Urearetics: a small molecule screen yields nanomolar potency inhibitors of urea transporter ut-b". *The FASEB Journal*, 21(2):551–563, **2007**.

[100] J. J. Doran, "Tissue distribution of UT-A and UT-B mRNA and protein in rat". *AJP: Regulatory, Integrative and Comparative Physiology*, 290(5):R1446–R1459, **2005**.

[101] M. A. Blount *et al.*, "Forskolin stimulates phosphorylation and membrane accumulation of UT-A3". *AJP: Renal Physiology*, 293(4):F1308–F1313, **2007**.

[102] J. D. Hoffert, "Quantitative phosphoproteomics of vasopressin-sensitive renal cells: Regulation of aquaporin-2 phosphorylation at two sites". *Proceedings of the National Academy of Sciences*, 103(18):7159–7164, **2006**.

[103] C. Zhang, J. M. Sands and J. D. Klein, "Vasopressin rapidly increases phosphorylation of UT-A1 urea transporter in rat IMCDs through PKA". *American journal of physiology Renal physiology*, 282(1):F85–90, **2002**.

[104] L. King and P. Agre, "From structure to disease: The evolving tale of aquaporin biology". *Nature Reviews Molecular Cell Biology*, 5:687–698, **2004**.

[105] D. Marx, "Proton transfer 200 years after von grotthuss: Insights from ab initio simulations". *ChemPhysChem*, 7(9):1848–1870, **2006**.

[106] D. Fu *et al.*, "Structure of a glycerol conducting channel and the basis for its selectivity". *Science*, 290:481–486, **2000**.

[107] D. F. Savage *et al.*, "Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z". *PLoS Biology*, 1:e72, **2003**.

[108] J. K. Lee *et al.*, "Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 Å". *Proceedings of the National Academy of Sciences, USA*, 102(52):18932 – 18937, **2005**.

[109] G. Fischer *et al.*, "Crystal structure of a yeast aquaporin at 1.15 Å reveals a novel gating mechanism". *PLoS Biology*, 7:e1000130, **2009**.

[110] Z. E. R. Newby *et al.*, "Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite Plasmodium falciparum". *Nature Structural & Molecular Biology*, 15(6):619–625, **2008**.

[111] S. Törnroth-Horsefield *et al.*, "Structural mechanism of plant aquaporin gating". *Nature*, 439:688–694, **2006**.

[112] H. Sui *et al.*, "Structural basis of water-specific transport through the AQP1 water channel". *Nature*, 414:872–878, **2001**.

[113] S. C. Harrison, "Whither structural biology?" *Nature Structural & Molecular Biology*, 11:12–15, **2004**.

[114] T. Gonen *et al.*, "Aquaporin-0 membrane junctions reveal the structure of a closed water pore". *Nature*, 429:193–197, **2004**.

[115] K. Murata *et al.*, "Structural determinants of water permeation through aquaporin-1". *Nature*, 407:599–605, **2000**.

[116] R. Horsefield *et al.*, "High-resolution X-ray structure of human aquaporin 5". *Proceedings of the National Academy of Sciences, USA*, 105:13327–32, **2008**.

[117] J. D. Ho *et al.*, "Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance". *Proceedings of the National Academy of Sciences, USA*, 106:7437–442, **2009**.

[118] Y. Wang, K. Schulten and E. Tajkhorshid, "What makes an aquaporin a glycerol channel: A comparative study of AqpZ and GlpF". *Structure*, 13:1107–1118, **2005**.

[119] E. Tajkhorshid *et al.*, "Control of the selectivity of the aquaporin water channel family by global orientational tuning". *Science*, 296:525–530, **2002**.

[120] B. L. de Groot and H. Grubmüller, "Water permeation across biological membranes: Mechanism and dynamics of aquaporin-1 and GlpF". *Science*, 294(5550):2353–2357, **2001**.

[121] B. Ilan *et al.*, "The mechanism of proton exclusion in aquaporin channels". *PROTEINS: Structure, Function, and Bioinformatics*, 55:223–228, **2004**.

[122] N. Chakrabarti *et al.*, "Molecular basis of proton blockage in aquaporins". *Structure*, 12:65–74, **2004**.

[123] B. L. de Groot *et al.*, "The mechanism of proton exclusion in the aquaporin-1 water channel". *Journal of Molecular Biology*, 333(2):279–293, **2003**.

[124] M. Ø. Jensen, E. Tajkhorshid and K. Schulten, "Electrostatic tuning of permeation and selectivity in aquaporin water channels". *Biophysical Journal*, 85:2884–2899, **2003**.

[125] B. L. de Groot and H. Grubmüller, "The dynamics and energetics of water permeation and proton exclusion in aquaporins". *Current Opinion in Structural Biology*, 15:1–8, **2005**.

[126] A. Burykin and A. Warshel, "What really prevents proton transport through aquaporin? Charge self-energy versus proton wire proposals". *Biophysical Journal*, 85(6):3696–3706, **2003**.

[127] B. Wu *et al.*, "Concerted action of two cation filters in the aquaporin water channel". *EMBO Journal*, 28(15):2188–2194, **2009**.

[128] D. Wree *et al.*, "Requirement for asparagine in the aquaporin NPA sequence signature motifs for cation exclusion: Role of asparagine in the aquaporin NPA motif". *FEBS Journal*, 278(5):740–748, **2011**.

[129] E. Beitz *et al.*, "Point mutations in the aromatic/arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia, and protons". *Proceedings of the National Academy of Sciences, USA*, 103:269–74, **2006**.

[130] H. Li *et al.*, "Enhancement of proton conductance by mutations of the selectivity filter of aquaporin-1". *Journal of Molecular Biology*, 407(4):607–620, **2011**, PMID: 21277313.

[131] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, New York, **1987**.

[132] D. E. Shaw *et al.*, "Millisecond-scale molecular dynamics simulations on anton", in "Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis", SC '09, 39:1–39:11, ACM, New York, NY, USA, **2009**.

[133] D. Frenkel and B. Smit, *Understanding Molecular Simulation From Algorithms to Applications*, Academic Press, California, **2002**.

[134] B. Roux, "The calculation of the potential of mean force using computer simulations". *Computer Physics Communications*, 91:275–282, **1995**.

[135] A. Barducci, G. Bussi and M. Parrinello, "Well-tempered metadynamics: a smoothly converging and tunable free-energy method". *Physical Review Letters*, 100:020603, **2008**.

[136] B. Isralewitz, M. Gao and K. Schulten, "Steered molecular dynamics and mechanical functions of proteins". *Current Opinion in Structural Biology*, 11:224–230, **2001**.

[137] J. Schlitter *et al.*, "Targeted molecular dynamics simulation of conformational change — application to the T $\leftrightarrow$ R transition in insulin". *Molecular Simulation*, 10(2–6):291–308, **1993**.

[138] J. Hénin and C. Chipot, "Overcoming free energy barriers using unconstrained molecular dynamics simulations". *Journal of Chemical Physics*, 121(7):2904–2914, **2004**.

[139] Y. Wang *et al.*, "Implementation of accelerated molecular dynamics in NAMD". *Computational Science & Discovery*, 4:015002, **2011**, (11 pages).

[140] M. Moradi and E. Tajkhorshid, "Mechanistic picture for conformational transition of a membrane transporter at atomic resolution". *Proceedings of the National Academy of Sciences, USA*, **2013**, in press.

[141] M. Moradi and E. Tajkhorshid, "Driven metadynamics: Reconstructing equilibrium free energies from driven adaptive-bias simulations." *Journal of Physical Chemistry Letters*, 4:1882, **2013**.

[142] S. Patel and C. L. Brooks, III, "CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations". *Journal of Computational Chemistry*, 25:1–15, **2004**.

[143] S. Patel, A. D. Mackerell, Jr. and C. L. Brooks, III, "CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model". *Journal of Computational Chemistry*, 25:1504–1514, **2004**.

[144] G. Lamoureux *et al.*, "A polarizable model of water for molecular dynamics simulations of biomolecules". *Chemical Physics Letters*, 418(1?3):245 – 249, **2006**.

[145] P. E. M. Lopes, B. Roux and A. D. MacKerell, Jr, "Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications". *Theoretica Chimica Acta*, 124:11–28, **2009**.

[146] E. Lindahl and M. S. P. Sansom, "Membrane proteins: molecular dynamics simulations". *Current Opinion in Structural Biology*, 18:425–431, **2008**.

[147] J. L. Klepeis *et al.*, "Long-timescale molecular dynamics simulations of protein structure and function". *Current Opinion in Structural Biology*, 19:120–127, **2009**.

[148] G. R. Bowman, V. A. Voelz and V. S. Pande, "Atomistic folding simulations of the five-helix bundle protein λ_{6-85}". *Journal of the American Chemical Society*, 133:664–667, **2011**.

[149] K. Lindorff-Larsen *et al.*, "How fast-folding proteins fold". *Science*, 334:517–520, **2011**.

[150] M. Ø. Jensen *et al.*, "Mechanism of voltage gating in potassium channels." *Science*, 336:229–233, **2012**.

[151] G. M. Torrie and J. P. Valleau, "Monte carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid". *Chemical Physics Letters*, 28:578–581, **1974**.

[152] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling". *Journal of Computational Physics*, 23:187–199, **1977**.

[153] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding". *Chemical Physics Letters*, 314:141–151, **1999**.

[154] Y. Sugita, A. Kitao and Y. Okamoto, "Multidimensional replica-exchange method for free-energy calculations". *Journal of Chemical Physics*, 113(15):6042–6051, **2000**.

[155] S. Park, T. Kim and W. Im, "Transmembrane helix assembly by window exchnage umbrella sampling". *Physical Review Letters*, 108:108102–108105, **2013**.

[156] S. Park and W. Im, "Two dimensional window exchange umbrella sampling for transmembrane helix assembly". *Journal of Chemical Theory and Computation*, 9:13–17, **2013**.

[157] M. Souaille and B. Roux, "Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations". *Computer Physics Communications*, 135:40–57, **2001**.

[158] C. H. Bennett, "Efficient estimation of free energy differences from Monte Carlo data". *Journal of Computational Physics*, 22:245–268, **1976**.

[159] C. Bartels, "Analyzing biased Monte Carlo and molecular dynamics simulations". *Chemical Physics Letters*, 331:446–454, **2000**.

[160] M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states". *Journal of Chemical Physics*, 129(12):124105, **2008**.

[161] A. D. MacKerell, Jr., M. Feig and C. L. Brooks, III, "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations". *Journal of Computational Chemistry*, 25(11):1400–1415, **2004**.

[162] J. B. Klauda *et al.*, "Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types". *Journal of Physical Chemistry B*, 114(23):7830–7843, **2010**.

[163] J. C. Phillips *et al.*, "Scalable molecular dynamics with NAMD". *Journal of Computational Chemistry*, 26:1781–1802, **2005**.

[164] W. Humphrey, A. Dalke and K. Schulten, "VMD – Visual Molecular Dynamics". *Journal of Molecular Graphics*, 14:33–38, **1996**.

[165] G. Enkavi and E. Tajkhorshid, "Simulation of spontaneous substrate binding revealing the binding pathway and mechanism and initial conformational response of GlpT". *Biochemistry*, 49:1105–1114, **2010**.

[166] M. J. Lemieux *et al.*, "Three-dimensional crystallization of the *Escherichia coli* glycerol-3-phosphate transporter: A member of the major facilitator superfamily". *Protein Science*, 12(12):2748–2756, **2003**.

[167] M. Auer *et al.*, "High-yield expression and functional analysis of *Escherichia coli* glycerol-3-phosphate transporter". *Biochemistry*, 40:6628–6635, **2001**.

[168] M. C. Fann *et al.*, "Identification of two essential arginine residues in uhpt, the sugar phosphate antiporter of *Escherichia coli*". *Journal of Membrane Biology*, 164(2):187–195, **1998**.

[169] M. Lemieux, Y. Huang and D. Wang, "The structural basis of substrate translocation by the glycerol-3-phosphate transporter: a member of the major facilitator superfamily". *Current Opinion in Structural Biology*, 14(4):405–412, **2004**.

[170] C. J. Law *et al.*, "Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT." *Journal of Molecular Biology*, 378:828–839, **2008**.

[171] I. F. Tsigelny *et al.*, "Modelling of glycerol-3-phosphate transporter suggests a potential 'tilt' mechanism involved in its function". *Journal of Bioinformatics and Computational Biology*, 6:885–904, **2008**.

[172] "The RCSB Protein Data Bank", http://www.rcsb.org/pdb.

[173] S. C. Lovell *et al.*, "Structure validation by C_α geometry: ϕ, ψ, and C_β deviation". *PROTEINS: Structure, Function, and Genetics*, 50:437–450, **2003**.

[174] "MolProbity", http://molprobity.biochem.duke.edu/.

[175] L. Zhang and J. Hermans, "Hydrophilicity of cavities in proteins". *PROTEINS: Structure, Function, and Genetics*, 24:433–438, **1996**.

[176] H. Grubmüller, "Solvate 1.0", **1996**.

[177] "Orientations of Proteins in Membranes (OPM) database", http://opm.phar.umich.edu/.

[178] M.-C. Fann and P. C. Maloney, "Functional symmetry of UhpT, the sugar phosphate transporter of *Escherichia coli*". *Journal of Biological Chemistry*, 273:33735–33740, **1998**.

[179] W. L. Jorgensen *et al.*, "Comparison of simple potential functions for simulating liquid water". *Journal of Chemical Physics*, 79(2):926–935, **1983**.

[180] G. J. Martyna, D. J. Tobias and M. L. Klein, "Constant pressure molecular dynamics algorithms". *Journal of Chemical Physics*, 101(5):4177–4189, **1994**.

[181] S. E. Feller *et al.*, "Constant pressure molecular dynamics simulation: The Langevin piston method". *Journal of Chemical Physics*, 103(11):4613–4621, **1995**.

[182] T. Darden, D. York and L. G. Pedersen, "Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems". *Journal of Chemical Physics*, 98(12):10089–10092, **1993**.

[183] S. V. Ambudkar, T. J. Larson and P. C. Maloney, "Reconstitution of sugar phosphate transport systems of *Escherichia coli*". *Journal of Biological Chemistry*, 261(20):9083–9086, **1986**.

[184] P. C. Maloney *et al.*, "Anion-exchange mechanisms in bacteria." *Microbiology and Molecular Biology Reviews*, 54(1):1–17, **1990**.

[185] O. S. Smart, J. M. Goodfellow and B. A. Wallace, "The pore dimensions of Gramicidin A". *Biophysical Journal*, 65(6):2455–2460, **1993**.

[186] R. S. G. D'rozario and M. S. P. Sansom, "Helix dynamics in a membrane transport protein: comparative simulations of the glycerol-3-phosphate transporter and its constituent helices". *Molecular Membrane Biology*, 25:571–573, **2008**.

[187] E. A. Johnson, C. Tanford and J. A. Reynolds, "Variable stoichiometry in active ion transport: theoretical analysis of physiological consequences". *Proceedings of the National Academy of Sciences, USA*, 82(16):5352–5356, **1985**.

[188] Y. Wang and E. Tajkhorshid, "Electrostatic funneling of substrate in mitochondrial inner membrane carriers". *Proceedings of the National Academy of Sciences, USA*, 105:9598–9603, **2008**.

[189] Y. Yin *et al.*, "Structure of the multidrug transporter EmrD from *Escherichia coli*". *Science*, 312(5774):741–744, **2006**.

[190] O. Mirza *et al.*, "Structural evidence for induced fit and a mechanism for sugar/H^+ symport in LacY". *EMBO Journal*, 25:1177–1183, **2006**.

[191] L. Guan *et al.*, "Structural determination of wild-type lactose permease". *Proceedings of the National Academy of Sciences*, 104(39):15294–15298, **2007**.

[192] S. Newstead *et al.*, "Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2." *EMBO Journal*, 30(2):417–426, **2011**.

[193] J. Abramson *et al.*, "Structure and mechanism of the lactose permease of escherichia coli." *Science*, 301(5633):610–615, **2003**.

[194] M. L. Mancianti and M. Herlyn, "Tumor progression in melanoma: the biology of epidermal melanocytes in vitro". *Carcinogenesis; a comprehensive survey*, 11:369–386, **1989**, PMID: 2646021.

[195] M. L. Mancianti and M. Herlyn, "Tumor progression in melanoma: the biology of epidermal melanocytes in vitro". *Carcinogenesis; a comprehensive survey*, 11:369–386, **1989**, PMID: 2646021.

[196] J. A. Hall, "Transmembrane segment 11 of UhpT, the sugar phosphate carrier of escherichia coli, is an alpha -helix that carries determinants of substrate selectivity". *Journal of Biological Chemistry*, 276(27):25107–25113, **2001**.

[197] M. I. Arbuckle *et al.*, "Structure-function analysis of liver-type (GLUT2) and brain-type (GLUT3) glucose transporters: expression of chimeric transporters in xenopus oocytes suggests an important role for putative transmembrane helix 7 in determining substrate selectivity". *Biochemistry*, 35(51):16519–16527, **1996**, PMID: 8987985.

[198] M. S. Collett, E. Erikson and R. L. Erikson, "Structural analysis of the avian sarcoma virus transforming protein: sites of phosphorylation". *Journal of virology*, 29(2):770–781, **1979**, PMID: 219258.

[199] R. T. Schweizer, "Mask wiggling as a potential cause of wound contamination". *Lancet*, 2(7995):1129–1130, **1976**, PMID: 62960.

[200] J. A. Hall, M. C. Fann and P. C. Maloney, "Altered substrate selectivity in a mutant of an intrahelical salt bridge in UhpT, the sugar phosphate carrier of escherichia coli". *The Journal of biological chemistry*, 274(10):6148–6153, **1999**, PMID: 10037698.

[201] J. A. Hall and P. C. Maloney, "Altered oxyanion selectivity in mutants of UhpT, the pi-linked sugar phosphate carrier of escherichia coli". *The Journal of biological chemistry*, 280(5):3376–3381, **2005**, PMID: 15556940.

[202] L. Guan, "Changing the lactose permease of escherichia coli into a galactose-specific symporter". *Proceedings of the National Academy of Sciences*, 99(10):6613–6618, **2002**.

[203] L. Guan, "Binding affinity of lactose permease is not altered by the h+ electrochemical gradient". *Proceedings of the National Academy of Sciences*, 101(33):12148–12152, **2004**.

[204] J. Wu and H. R. Kaback, "Cysteine 148 in the lactose permease of escherichia coli is a component of a substrate binding site. 2. site-directed fluorescence studies". *Biochemistry*, 33(40):12166–12171, **1994**.

[205] R. J. Naftalin, N. Green and P. Cunningham, "Lactose permease h+-lactose symporter: Mechanical switch or brownian ratchet?" *Biophysical Journal*, 92(10):3474–3491, **2007**.

[206] I. Kwaw *et al.*, "Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli* : Helices IV and v that contain the major determinants for substrate binding ". *Biochemistry*, 40(35):10491–10499, **2001**.

[207] L. Guan and H. R. Kaback, "Lessons from lactose permease". *Annual Review of Biophysics and Biomolecular Structure*, 35(1):67–91, **2006**.

[208] R. Vink *et al.*, "Estimation of h+ to adenosine 5'-triphosphate stoichiometry of escherichia coli ATP synthase using 31P NMR". *Biochemistry*, 23(16):3667–3675, **1984**, PMID: 6089877.

[209] J. Li *et al.*, "Transient formation of water-conducting states in membrane transporters". *Proceedings of the National Academy of Sciences, USA*, 110(19):7696–7701, **2013**.

[210] T. Zeuthen, "Water-transporting proteins." *Journal of Membrane Biology*, 234(2):57–73, **2010**.

[211] O. Andrini *et al.*, "The leak mode of type II Na^+-P_i cotransporters". *Channels*, 2:346–357, **2008**.

[212] A. Conde *et al.*, "Transporters, channels, or simple diffusion? dogmas, atypical roles and complexity in transport systems." *Int J Biochem Cell Biol*, 42:857–868, **2010**.

[213] T. Zeuthen and W. D. Stein, "Cotransport of salt and water in membrane proteins: Membrane proteins as osmotic engines." *Journal of Membrane Biology*, 137:179–195, **1994**.

[214] P. P. Duquette, P. Bissonnette and J. Y. Lapointe, "Local osmotic gradients drive the water flux associated with Na^+/glucose cotransport." *Proceedings of the National Academy of Sciences, USA*, 98(7):3796–3801, **2001**.

[215] M. P. Gagnon *et al.*, "Glucose accumulation can account for the initial water flux triggered by Na^+/glucose cotransport." *Biophysical Journal*, 86:125–133, **2004**.

[216] R. J. Naftalin, "Osmotic water transport with glucose in GLUT2 and SGLT". *Biophysical Journal*, 94:3912–3923, **2008**.

[217] A. Meinild *et al.*, "The human Na^+-glucose cotransporter is a molecular water pump." *Journal of Physiology*, 508(Pt 1):15–21, **1998**.

[218] T. Zeuthen, B. Belhage and E. Zeuthen, "Water transport by Na^+-coupled cotransporters of glucose (SGLT1) and of iodide (NIS). the dependence of substrate size studied at high resolution." *Journal of Physiology*, 570(Pt 3):485–499, **2006**.

[219] J. Fischbarg *et al.*, "Glucose transporters serve as water channels." *Proceedings of the National Academy of Sciences, USA*, 87(8):3244–3247, **1990**.

[220] T. Zeuthen, E. Zeuthen and N. Macaulay, "Water transport by GLUT2 expressed in *Xenopus laevis* oocytes." *Journal of Physiology*, 579(2):345–361, **2007**.

[221] J. C. Gordon *et al.*, "H++: a server for estimating pK_as and adding missing hydrogens to macromolecules." *Nucleic Acids Research*, 33(Web Server issue):W368–371, **2005**, http://biophysics.cs.vt.edu/H++.

[222] L. J. DeFelice and T. Goswami, "Transporters as channels". *Annual Review of Physiology*, 69:87–112, **2007**.

[223] D. D. Loo *et al.*, "Passive water and ion transport by cotransporters." *Journal of Physiology*, 518(1):195–202, **1999**.

[224] M. Santacroce, M. Castagna and V. F. Sacchi, "Passive water permeability of some wild type and mutagenized amino acid cotransporters of the SLC6/NSS family expressed in *Xenopus laevis* oocytes." *Comp Biochem Physiol A Mol Integr Physiol*, 156(4):509–517, **2010**.

[225] N. MacAulay *et al.*, "Passive water and urea permeability of a human Na^+-glutamate cotransporter expressed in *Xenopus* oocytes". *Journal of Physiology*, 542:817–828, **2002**.

[226] P.-C. Wen *et al.*, "Molecular mechanisms of active transport across the cellular membrane", in P. Biggin and M. Sansom (Editors), "Molecular Simulations and Biomembranes: From Biophysics to Function", chapter 10, 248–286, Royal Society of Chemistry, **2010**.

[227] S. A. Shaikh *et al.*, "Visualizing functional motions of membrane transporters with molecular dynamics simulations." *Biochemistry*, 52(4):569–587, **2013**.

[228] Z. Huang *et al.*, "Membrane transporters: Molecular machines coupling cellular energy to vectorial transport across the membrane", in B. Roux (Editor), "Molecular Machines", chapter 9, 151–182, World Scientific Publishing, Singapore, **2011**.

[229] M. Tirion, "Large amplitude elastic motions in proteins from a single-parameter atomic analysis". *Physical Review Letters*, 77:1905–1908, **1996**.

[230] A. R. Atilgan *et al.*, "Anisotropy of fluctuation dynamics of proteins with an elastic network model". *Biophysical Journal*, 80:505–515, **2001**.

[231] B. Erman, "The Gaussian Network Model: Precise Predictions of Residue Fluctuations and Application to Binding Problems". *Biophysical Journal*, 91(10):3589–3599, **2006**.

[232] E. Eyal, L. W. Yang and I. Bahar, "Anisotropic network model: systematic evaluation and a new web interface". *Bioinformatics*, 22:2619–2627, **2006**.

[233] B. MacIver *et al.*, "Functional characterization of mouse urea transporters UT-A2 and UT-A3 expressed in purified xenopus laevis oocyte plasma membranes". *AJP: Renal Physiology*, 294(4):F956–F964, **2008**.

[234] L. M. Mannuzzu, M. M. Moronne and R. I. Macey, "Estimate of the number of urea transport sites in erythrocyte ghosts using a hydrophobic mercurial". *The Journal of membrane biology*, 133(1):85–97, **1993**.

[235] S. Raunser *et al.*, "Oligomeric structure and functional characterization of the urea transporter from actinobacillus pleuropneumoniae". *Journal of Molecular Biology*, 387(3):619–627, **2009**.

[236] S. Kumar *et al.*, "The weighted histogram analysis method for free-energy calculations on biomolecules". *Journal of Computational Chemistry*, 13:1011–1021, **1992**.

[237] A. Caflisch and M. Karplus, "Structural details of urea binding to barnase: a molecular dynamics analysis". *Structure*, 7(5):477–488, **1999**, PMID: 10378267.

[238] G. Rousselet, P. Ripoche and P. Bailly, "Tandem sequence repeats in urea transporters: identification of an urea transporter signature sequence". *The American journal of physiology*, 270(3 Pt 2):F554–555, **1996**.

[239] R. Minocha, K. Studley and J. Saier, Milton H, "The urea transporter (UT) family: bioinformatic analyses leading to structural, functional, and evolutionary predictions". *Receptors & channels*, 9(6):345–352, **2003**.

[240] S. Khademi *et al.*, "Mechanism of ammonia transport by Amt/MEP/Rh: Structure of AmtB at 1.35 Å". *Science*, 1587–1594, **2004**.

[241] L. Zheng, "The mechanism of ammonia transport based on the crystal structure of AmtB of escherichia coli". *Proceedings of the National Academy of Sciences*, 101(49):17090–17095, **2004**.

[242] B. Maciver *et al.*, "Functional characterization of mouse urea transporters UT-a2 and UT-a3 expressed in purified xenopus laevis oocyte plasma membranes." *American Journal of Physiology – Renal Physiology*, 294(4), **2008**.

[243] D. Zhao *et al.*, "Comparative transport efficiencies of urea analogues through urea transporter UT-B". *Biochimica et biophysica acta*, 1768(7):1815–1821, **2007**.

[244] B. Yang, "Urea transporter UT3 functions as an efficient water channel. Direct evidence for a common water/urea pathway". *Journal of Biological Chemistry*, 273(16):9369–9372, **1998**.

[245] J. D. Klein, "Vasopressin increases plasma membrane accumulation of urea transporter UT-A1 in rat inner medullary collecting ducts". *Journal of the American Society of Nephrology*, 17(10):2680–2686, **2006**.

[246] G. S. Stewart *et al.*, "Vasopressin regulation of the renal UT-A3 urea transporter". *AJP: Renal Physiology*, 296(3):F642–F648, **2008**.

[247] J. P. Kokko and F. C. Rector, "Countercurrent multiplication system without active transport in inner medulla". *Kidney International*, 2(4):214–223, **1972**.

[248] A. T. Layton, "Role of UTB urea transporters in the urine concentrating mechanism of the rat kidney". *Bulletin of Mathematical Biology*, 69(3):887–929, **2007**.

[249] J. L. Stephenson, "Concentration of urine in a central core model of the renal counterflow system". *Kidney international*, 2(2):85–94, **1972**, PMID: 4671532.

[250] T. L. Pallone *et al.*, "Evidence that aquaporin-1 mediates NaCl-induced water flux across descending vasa recta". *The American journal of physiology*, 272(5 Pt 2):F587–596, **1997**, PMID: 9176368.

[251] R. R. Geyer *et al.*, "Movement of NH_3 through the human urea transporter b: a new gas channel". *American Journal of Physiology – Renal Physiology*, 304(12):F1447–F1457, **2013**.

[252] G. Cooper and W. Boron, "The CO2 permeability of the AQP1 water channel, expressed in xenopus oocytes.", volume 8, A0074–A0074, WILLIAMS & WILKINS 351 WEST CAMDEN ST, BALTIMORE, MD 21201-2436, **1997**.

[253] G. J. Cooper and W. F. Boron, "Effect of PCMBS on CO2 permeability of xenopus oocytes expressing aquaporin 1 or its C189S mutant". *The American journal of physiology*, 275(6 Pt 1):C1481–1486, **1998**.

[254] V. Endeward *et al.*, "Evidence that aquaporin 1 is a major pathway for CO2 transport across the human erythrocyte membrane". *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 20(12):1974–1981, **2006**.

[255] R. E. Forster *et al.*, "The effect of 4,4'-diisothiocyanato-stilbene-2,2'-disulfonate on CO2 permeability of the red blood cell membrane". *Proceedings of the National Academy of Sciences of the United States of America*, 95(26):15815–15820, **1998**.

[256] R. Musa-Aziz *et al.*, "Relative CO2/NH3 selectivities of AQP1, AQP4, AQP5, AmtB, and RhAG". *Proceedings of the National Academy of Sciences of the United States of America*, 106(13):5406–5411, **2009**.

[257] N. L. Nakhoul *et al.*, "Effect of expressing the water channel aquaporin-1 on the CO2 permeability of xenopus oocytes". *The American journal of physiology*, 274(2 Pt 1):C543–548, **1998**.

[258] A. M. Marini *et al.*, "The human rhesus-associated RhAG protein and a kidney homologue promote ammonium transport in yeast". *Nature genetics*, 26(3):341–344, **2000**.

[259] N. L. Nakhoul *et al.*, "Transport of NH(3)/NH in oocytes expressing aquaporin-1". *American journal of physiology Renal physiology*, 281(2):F255–263, **2001**.

[260] P. Ripoche *et al.*, "Role of RhAG and AQP1 in NH3 and CO2 gas transport in red cell ghosts: a stopped-flow analysis". *Transfusion clinique et biologique: journal de la Socit franaise de transfusion sanguine*, 13(1-2):117–122, **2006**.

[261] W. F. Boron, "Sharpey-schafer lecture: Gas channels". *Experimental Physiology*, 95(12):1107–1130, **2010**.

[262] M. Herrera and J. L. Garvin, "Novel role of AQP-1 in NO-dependent vasorelaxation". *American journal of physiology Renal physiology*, 292(5):F1443–1451, **2007**, PMID: 17229677.

[263] M. Herrera, N. J. Hong and J. L. Garvin, "Aquaporin-1 transports NO across cell membranes". *Hypertension*, 48(1):157–164, **2006**.

[264] S. Waisbren *et al.*, "Unusual permeability properties of gastric gland cells". *Nature*, 368:332–335, **1994**.

[265] Y. Wang, Y. Z. Ohkubo and E. Tajkhorshid, "Gas conduction of lipid bilayers and membrane channels", in S. Feller (Editor), "Current Topics in Membranes: Computational Modeling of Membrane Bilayers", volume 60, chapter 12, 343–367, Elsevier, **2008**.

[266] Y. Wang and E. Tajkhorshid, "Nitric oxide conduction by the brain aquaporin AQP4". *PROTEINS: Structure, Function, and Bioinformatics*, 78:661–670, **2010**.

[267] S. Khademi, "Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 a". *Science*, 305(5690):1587–1594, **2004**.

[268] P. Ripoche, "Human rhesus-associated glycoprotein mediates facilitated transport of NH3 into red blood cells". *Proceedings of the National Academy of Sciences*, 101(49):17222–17227, **2004**.

[269] V. Endeward *et al.*, "RhAG protein of the rhesus complex is a CO2 channel in the human red cell membrane". *The FASEB Journal*, 22(1):64–73, **2007**.

[270] S. Jo *et al.*, "CHARMM-GUI: A web-based graphical user interface for CHARMM". *Journal of Computational Chemistry*, 29:1859–1865, **2008**.

[271] S. Jo. *et al.*, "CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes". *Biophysical Journal*, 97:50–58, **2009**.

[272] A. Wlodawer *et al.*, "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures: Protein crystallography for non-crystallographers". *FEBS Journal*, 275(1):1–21, **2008**.

[273] Y. Zhou *et al.*, "Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 Å resolution". *Nature*, 414:43–48, **2001**.

[274] J. H. Morais-Cabral, Y. Zhou and R. MacKinnon, "Energetic optimization of ion conduction rate by the K^+ selectivity filter". *Nature*, 414:37–41, **2001**.